



***Facultad  
de  
Ciencias***

**Detección y clasificación de incendios  
mediante métodos de aprendizaje  
automático sobre imágenes de Sentinel-2**

**(Fire detection and classification using  
automatic learning methods on Sentinel-2  
images)**

**Trabajo de Fin de Grado  
para acceder al**

**GRADO EN FÍSICA**

**Autor: Oliver Legarreta García**

**Director: Jesús Marco de Lucas**

**Co-Director: Daniel García Díaz**

**Junio - 2020**

## **Agradecimientos**

En primer lugar, agradecer a Jesús Marco por darme la oportunidad de realizar este trabajo de fin de grado. Y sobre todo, agradecer a Fernando Aguilar y Daniel García por su ayuda, paciencia, apoyo y el buen trato recibido desde el principio hasta el final de este trabajo. Han hecho que este proceso haya sido más cómodo, lo que todo estudiante desea. Por ello, siempre les estaré agradecido.

Mi interés en la física se debe, en gran medida, a dos personas. Por una parte, gracias a Alberto Aguayo por transmitirme su pasión por la física dentro y fuera de las clases. Por otra parte, gracias a Amaya Cobo por lo buena profesora, persona y física que es. Siempre diré que sin ella no hubiese empezado esta carrera.

En general, gracias a amigos y familiares. Y en concreto, gracias a mi madre por, sencillamente, todo.

Por último y de forma más especial, gracias a mi abuelo, responsable de todo lo bueno que me ha pasado en esta vida. Que en paz descanse.

## Resumen

La teledetección satelital se ha convertido en una de las fuentes de datos principales en detección y monitorización de incendios. En este proyecto se utilizan tres técnicas computacionales de clasificación de datos con los objetivos de detectar, delimitar y analizar la superficie afectada por varios incendios y su evolución con el tiempo. Para ello se emplean los índices de vegetación NBR y NDVI, y datos de los satélites Sentinel-2. Los resultados obtenidos son validados mediante parámetros estadísticos propios a la matriz de confusión, permitiendo determinar la eficacia de cada técnica.

**Palabras claves:** Teledetección, NBR, NDVI, Resta de Imágenes, K-Means, Regresión Logística, Matriz de Confusión

## Abstract

Remote sensing has become one of the main sources of data in fire detection and monitoring. In this project, three computerized data classification techniques are used with the aim of detecting, delimiting and analyzing the surface affected by several fires and their evolution over time. For this purpose, the NBR and NDVI vegetation indices and data from the Sentinel-2 satellites are used. The results obtained are validated using statistical parameters specific to the confusion matrix, allowing the effectiveness of each technique to be determined.

**Key words:** Remote Sensing, NBR, NDVI, Image Subtraction, K-Means, Logistic Regression, Confusion Matrix

# Índice

<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos y motivación del trabajo . . . . .	1
1.2. Los incendios forestales y sus efectos . . . . .	1
1.2.1. Efectos en la salud . . . . .	2
1.2.2. Efectos en la biodiversidad . . . . .	2
1.2.3. Efectos y relación con el cambio climático . . . . .	3
1.3. Áreas de estudio . . . . .	4
1.3.1. Encinedo, León . . . . .	5
1.3.2. Galicia . . . . .	5
<b>2. Teledetección Satelital</b>	<b>7</b>
2.1. Teledetección . . . . .	7
2.1.1. El programa Copernicus . . . . .	8
2.1.2. Principios físicos de la teledetección . . . . .	9
2.1.3. El espectro electromagnético . . . . .	10
2.1.4. Firma espectral . . . . .	11
2.2. Índices de Vegetación . . . . .	12
2.2.1. Índice NDVI . . . . .	12
2.2.2. Índice NBR . . . . .	13
<b>3. Metodología y Técnicas Experimentales</b>	<b>15</b>
3.1. Adquisición y procesamiento de datos . . . . .	15
3.2. Métodos empleados para la detección de los incendios . . . . .	15
3.2.1. Resta de imágenes . . . . .	16
3.2.2. Aprendizaje no supervisado: K-Means . . . . .	17
3.2.3. Aprendizaje supervisado: Regresión logística . . . . .	18
3.2.4. Parámetros de validación: la matriz de confusión . . . . .	19
3.3. Procedimiento Empleado . . . . .	21
3.3.1. Detección de los incendios . . . . .	21
3.3.2. Avance temporal de los incendios . . . . .	22
<b>4. Resultados y Análisis</b>	<b>24</b>
4.1. Detección de los incendios . . . . .	24



4.1.1.	Mascaras Originales: Encinedo . . . . .	24
4.1.2.	Máscaras Originales: Galicia . . . . .	25
4.1.3.	Resta de Imágenes: Encinedo . . . . .	26
4.1.4.	Resta de imágenes: Galicia . . . . .	28
4.1.5.	Método no supervisado, K-Means: Encinedo . . . . .	31
4.1.6.	Método no supervisado, K-Means: Galicia . . . . .	33
4.1.7.	Método supervisado, Logistic Regression: Encinedo . . . . .	34
4.1.8.	Método supervisado, Logistic Regression: Galicia . . . . .	37
4.2.	Análisis de resultados de detección de incendios . . . . .	39
4.3.	Evolución temporal de la superficie del incendio . . . . .	41
4.3.1.	Encinedo . . . . .	41
4.3.2.	Galicia . . . . .	43
4.4.	Análisis de resultados del avance temporal del incendio . . . . .	45
<b>5.</b>	<b>Conclusiones</b>	<b>47</b>

# Índice de figuras

1.	Perímetro incendio Encinedo . . . . .	5
2.	Incendios Galicia . . . . .	6
3.	Proceso Teledetección . . . . .	7
4.	Espectro electromagnético . . . . .	11
5.	Firma espectral . . . . .	11
6.	Ejemplo resta de imágenes . . . . .	16
7.	Ejemplo del método del codo utilizando los datos de los índices NBR y NDVI del incendio de Enciendo. . . . .	18
8.	Función sigmoide . . . . .	19
9.	Matriz de confusión . . . . .	20
10.	Máscaras originales Encinedo . . . . .	24
11.	Máscaras originales Galicia . . . . .	25
12.	Resta de imágenes Encinedo . . . . .	26
13.	MC resta de imágenes NBR Encinedo . . . . .	26
14.	Resta de imágenes NDVI Encinedo . . . . .	27
15.	Matriz de confusión para el modelo de resta de imágenes con datos del índice NDVI sobre el incendio de Encinedo. . . . .	28
16.	Resta de imágenes NBR Galicia . . . . .	29
17.	Matriz de confusión para el índice NBR de la resta de imágenes en Galicia. . . . .	29
18.	Resta de imágenes NDVI Galicia . . . . .	30
19.	Matriz de confusión para el índice NDVI de la resta de imágenes en Galicia. . . . .	30
20.	K-Means Encinedo . . . . .	32
21.	Matriz de confusión para el método K-Means en Encinedo. . . . .	32
22.	K-Means Galicia . . . . .	33
23.	Matriz de confusión para el método K-Means en Galicia. . . . .	34
24.	Método supervisado NBR Encinedo . . . . .	35
25.	Matriz de confusión para el índice NBR en el método supervisado. . . . .	35
26.	Método supervisado NDVI Encinedo . . . . .	36
27.	Matriz de confusión para el método supervisado en el incendio de Encinedo para el índice NDVI. . . . .	36
28.	Método supervisado NBR Galicia . . . . .	37
29.	Matriz de confusión para el índice NBR en el método supervisado en Galicia. . . . .	38

30.	Método supervisado NDVI Galicia . . . . .	38
31.	Matriz de confusión para el índice NDVI en el método supervisado en Galicia. . . . .	39
32.	Evolución temporal NBR Encinedo . . . . .	41
33.	Píxeles NBR evolución temporal Encinedo . . . . .	42
34.	Evolución temporal NDVI Encinedo . . . . .	42
35.	Píxeles evolución temporal NDVI Encinedo . . . . .	43
36.	Evolución temporal NBR Galicia . . . . .	43
37.	Píxeles evolución temporal dNBR Galicia . . . . .	44
38.	Evolución temporal NDVI Encinedo . . . . .	44
39.	Píxeles evolución temporal NDVI Galicia . . . . .	45
40.	Vegetación de Encinedo y Galicia según QBIF. . . . .	46

## Índice de tablas

1.	Bandas Sentinel-2 . . . . .	9
2.	Valores NDVI . . . . .	13
3.	Valores dNBR . . . . .	14
4.	Píxeles incendio máscaras originales Encinedo . . . . .	24
5.	Píxeles incendio máscaras originales Galicia . . . . .	25
6.	Parámetros validación resta de imágenes NBR Encinedo . . . . .	27
7.	Parámetros validación resta de imágenes NDVI Encinedo . . . . .	28
8.	Parámetros validación resta de imágenes NBR Galicia . . . . .	29
9.	Parámetros validación resta de imágenes NDVI Encinedo . . . . .	31
10.	Resultados K-Means Encinedo . . . . .	31
11.	Parámetros validación K-Means Encinedo . . . . .	32
12.	Resultados K-Means Galicia . . . . .	33
13.	Parámetros validación K-Means Galicia . . . . .	34
14.	Resultados método supervisado NBR Encinedo . . . . .	34
15.	Parámetros validación método supervisado NBR Encinedo . . . . .	35
16.	Parámetros validación método supervisado NDVI Encinedo . . . . .	36
17.	Método supervisado NBR Galicia . . . . .	37
18.	Parámetros validación método supervisado NBR Galicia . . . . .	38
19.	Parámetros validación método supervisado NDVI Galicia . . . . .	39

20.	Resumen de resultados de los tres métodos en Encinedo . . . . .	39
21.	Resumen de resultados de los tres métodos en Galicia . . . . .	40

# **1. Introducción**

## **1.1. Objetivos y motivación del trabajo**

Mediante la teledetección satelital se consigue una gran cantidad de datos que han de ser procesados y analizados para extraer información relevante y útil. Existen varias aplicaciones, pero es su uso en lo referente a incendios forestales el tratado en este proyecto. La combinación de técnicas de análisis de datos y la información recibida mediante teledetección permite entender, controlar y tomar decisiones, de forma rápida y eficaz, de cara a un incendio forestal. Por ello, en este trabajo se busca entender el proceso de teledetección, presentar y aplicar técnicas de clasificación de datos e interpretar los resultados obtenidos en los incendios de 2017 de Encinedo (León) y Galicia.

El primer y principal objetivo de este trabajo es la utilización de imágenes de satélites para detectar y delimitar dos zonas, de origen geográfico distinto, afectadas por un incendio. Para ello, se emplean tres técnicas: un método directo (resta de imágenes), un modelo de minería de datos (método no supervisado, K-Means) y un modelo de Machine Learning (método supervisado, Regresión Logística). El resultado de estas técnicas son imágenes que representan el área dañada tras un incendio. Los datos satelitales usados provienen de la misión Sentinel-2 del programa Copernicus de la ESA.

El segundo objetivo concierne el estudio y la observación de las zonas incendiadas con el paso del tiempo. Mediante el uso de una de las tres técnicas empleadas y los índices NBR y NDVI, se analiza la regeneración de la vegetación del entorno deteriorado por el incendio. Para ello, se busca dar una explicación, desde el punto de vista de la vegetación, del motivo por el que un área se regenera más rápido que otra.

La unión de ambos objetivos tiene como resultado una técnica capaz de detectar y delimitar un incendio así como monitorizar su evolución con el tiempo.

La teledetección satelital tiene como principales ventajas su cobertura consistente, periódica y de gran área, y su facilidad para proporcionar información de superficies remotas [1]. Además, con el paso del tiempo y cada vez más rápido, sus sistemas mejoran consiguiéndose mejores resultados. Descubrir y perfeccionar técnicas de análisis de datos provenientes de satélites, permite facilitar la detección y el control de los incendios, adquirir mejor información para estudios y por lo tanto, disminuir sus consecuencias. Y es esta la motivación principal del trabajo: encontrar técnicas que permitan disminuir las consecuencias de los incendios forestales.

## **1.2. Los incendios forestales y sus efectos**

Según la agencia estatal de meteorología (AEMET), un incendio forestal es un fuego de gran magnitud que se propaga sin control en algún terreno forestal, es de-

cir, conformado básicamente por árboles [2]. Se caracteriza por su amplia extensión, la velocidad con la que se puede extender desde su punto de origen y su capacidad para superar obstáculos. Se pueden clasificar según su tamaño, denominándose como *grandes incendios* aquellos que afectan a superficies iguales o mayores a 500 *ha*, los cuales son tratados en este trabajo.

Los incendios forestales tienen efectos de distinta índole. Sus consecuencias son por un lado inmediatas y por otro lado duraderas a lo largo del tiempo. Puede ser difícil valorar de forma completa y correcta los efectos. Para entenderlos, lo mejor es hablar de los más importantes: en la salud, en la biodiversidad y en el cambio climático.

### 1.2.1. Efectos en la salud

Se podría decir que la pérdida de vidas humana es la peor consecuencia de los incendios forestales. Aunque es cierto que las cifras de fallecidos son más bajas que las cifras de fallecidos por otras catástrofes, es importante tenerlas en cuenta. Entre 1980 y 2010 han fallecido 245 personas en España por causa directa de incendios forestales mientras que 1670 han resultado heridas [3]. Además, hay un importante número de personas que mueren (y no son contabilizadas) o son afectadas a posteriori por consecuencias indirectas generalmente de carácter ambiental, en concreto por los humos generados.

Cuando hablamos de humos generados en realidad lo hacemos de partículas en suspensión. Las partículas en suspensión (particulate matter, PM) son partículas que contaminan el aire. Diversos estudios [4] han demostrado que altos niveles de PM<sub>2,5</sub> (partículas más pequeñas que 2.5  $\mu\text{m}$ ) están relacionadas con efectos adversos en la salud, incluyendo problemas respiratorios, neumonía, problemas del corazón, obstrucción pulmonar y muerte prematura. Aproximadamente un tercio de las emisiones de PM<sub>2,5</sub> provienen de incendios forestales debido a la quema de biomasa que incluye sulfatos, nitratos, mercurio y polvo [5]. Los niveles de PM dependerán del tipo de biomasa quemada y de la intensidad del fuego. Se estima que la mortalidad diaria aumenta un 0.5 % por cada incremento de 10  $\frac{\mu\text{g}}{\text{m}^3}$  de PM<sub>10</sub> y que existe un aumento del 2 % del riesgo de neumonía y obstrucción pulmonar con este mismo incremento de PM<sub>10</sub>.

### 1.2.2. Efectos en la biodiversidad

El fuego es un elemento esencial y natural en el funcionamiento de numerosos ecosistemas forestales. Ha sido una de las principales fuerzas evolutivas que han moldeado la estructura, composición y distribución geográfica de los ecosistemas con coberturas vegetales del mundo. Además ha equilibrado el impacto sobre los componentes estructurales, dinámicas, interrelaciones y procesos de dichos ecosis-

temas [6]. Sin embargo, el cambio de la dinámica establecida entre la actuación humana y los incendios han dado lugar a una situación en la que estos son una amenaza importante para muchos bosques y la diversidad que contienen.

Modifican el volumen de biomasa, alteran el ciclo hidrológico e influyen en el comportamiento de las especies vegetales y animales. Uno de los efectos ecológicos más importantes de los incendios es la mayor probabilidad de que se produzcan nuevos episodios del mismo tipo en los años subsiguientes, al caer los árboles, lo que permite que la luz del sol reseque el bosque y produzca una acumulación de combustible con un aumento de especies susceptibles a los incendios, como las herbáceas inflamables [7].

Pueden impactar de distintas formas y en diferentes momentos a las especies vegetales presentes en una localidad. Siendo los efectos más importantes: la alteración de las condiciones microclimáticas necesarias para su reproducción y desarrollo, la reducción del tamaño de las poblaciones locales por muerte directa o diferida de sus individuos, la modificación de sus estructuras poblacionales, la disminución de sus diversidad genética, la generación de condiciones para el ingreso de nuevas especies competidoras y la alteración de las interacciones biológicas y ecológicas con otras especies no solo vegetales [8].

### **1.2.3. Efectos y relación con el cambio climático**

El cambio climático propicia la generación y propagación de incendios forestales y a su vez estos alimentan al cambio climático, de esta forma tanto cambio climático como incendios forestales se agravan mutuamente con el tiempo.

La mayor parte de la energía que llega a la Tierra proveniente del Sol, lo hace en forma de radiación electromagnética. La radiación solar entrante ha de ser equivalente a la radiación emitida y reflejada por la Tierra, tiene que existir un equilibrio. Un cambio en este equilibrio genera variaciones climáticas en la Tierra. Parte de la radiación solar recibida permanece en nuestro planeta, permitiendo condiciones de temperatura agradables para la vida, gracias al efecto invernadero. Se denomina efecto invernadero al fenómeno por el cual determinados gases, que son componentes de la atmósfera planetaria, retienen parte de la energía que el suelo emite y refleja tras haber sido calentado por la radiación solar [9]. Ahora bien, el incremento de gases de efecto invernadero puede provocar la acumulación excesiva de radiación solar, lo que supone el aumento de las temperaturas terrestres. Los gases de efecto invernadero absorben y emiten radiación en determinadas longitudes de ondas del espectro de radiación infrarroja emitido por la superficie de la Tierra, la atmósfera, y las nubes. Entre ellos destaca el  $\text{CO}_2$ . ¿Cómo se generan mayores cantidades de  $\text{CO}_2$ ? Principalmente con la quema de combustibles fósiles. Estudios estiman que el 25 % de las emisiones globales anuales de los combustibles fósiles se

deben a incendios forestales [10]. Otros estudios [11] [12] señalan que los incendios forestales afectan al cambio climático de dos formas más: el carbón negro que se deposita en el Ártico, ya que disminuye el albedo del hielo y consecuentemente el porcentaje de radiación reflejada, y la capacidad de reducción de absorción de CO<sub>2</sub> que se pierde debido a las quemaduras de los incendios forestales.

El riesgo de incendio forestal depende de varios factores, entre ellos la temperatura, la humedad del suelo y la presencia de árboles, arbustos y otros posibles combustibles. Todos estos factores se ven afectados por el cambio climático en mayor o menor medida. Según el Panel Intergubernamental sobre Cambio Climático (IPCC), los ecosistemas mediterráneos se encuentran entre los más vulnerables del mundo y sufrirán impactos muy severos si no se reduce drásticamente la emisión de gases de efecto invernadero [13].

Los escenarios regionalizados para España indican un aumento de las temperaturas para finales de siglo, 4°C en invierno y 6°C en verano, los periodos de fuertes sequías y olas de calor extremas serán más frecuentes y duraderos, el número de días con precipitaciones disminuirá un 25 % al año y la duración de la temporada de incendios será mayor. Si el aumento de temperatura global supera los 2°C, los días de riesgo de incendio se incrementarán en hasta seis semanas al año. Según un informe del Instituto de Física de Cantabria [14], antes de 2075, las áreas quemadas por incendios se triplicarán en la península ibérica respecto a las que arden en la actualidad, como consecuencia del calentamiento global.

### **1.3. Áreas de estudio**

Debido al impacto de los incendios forestales en la península Ibérica, se ha decidido escoger dos incendios representativos de gran magnitud en este territorio. Primero se ha valorado el número de hectáreas arrasadas. También se ha tenido en cuenta la disponibilidad de archivos. Sentinel-2, en la sección 3 se proporcionan detalles de esta misión satelital, comienza en 2015. Por ello, se han considerado los incendios forestales en España desde ese año. De esta forma, el primer incendio escogido es el de Encinedo (León), que está considerado el mayor incendio en España desde 2015. El segundo incendio no es uno en concreto, son varios. En octubre de 2017 una serie de fuertes incendios arrasaron miles de hectáreas en Galicia. La resolución de Sentinel-2 permite abarcarlos todos en una misma imagen. Entre ellos destacan el de Pontearreas (Pontevedra) y Melón (Ourense), dos de los incendios más fuertes desde 2015. Por último, se ha tenido en cuenta el lugar donde los incendios se han originado para poder comparar la evolución de estos en función de su vegetación.



### 1.3.1. Encinedo, León

Encinedo pertenece a la comarca de La Cabrera, la cual se encuentra en el extremo suroccidental de León, limitando con las provincias de Zamora y Ourense. Su terreno es montañoso caracterizado por su relieve acusado, cuya superficie es de 1496 km<sup>2</sup>. El territorio consta principalmente de suelos ácidos y pobres donde la vegetación está compuesta en mayor medida por matorrales. El clima es oceánico-continental, caracterizado por inviernos fríos con alto nivel de precipitaciones y veranos cálidos con fuertes sequías. El incendio de Encinedo se originó el 21 de agosto de 2017 y arrasó con 9820 *ha*. Fue el incendio forestal que mayor superficie devastó de 2017 [15].

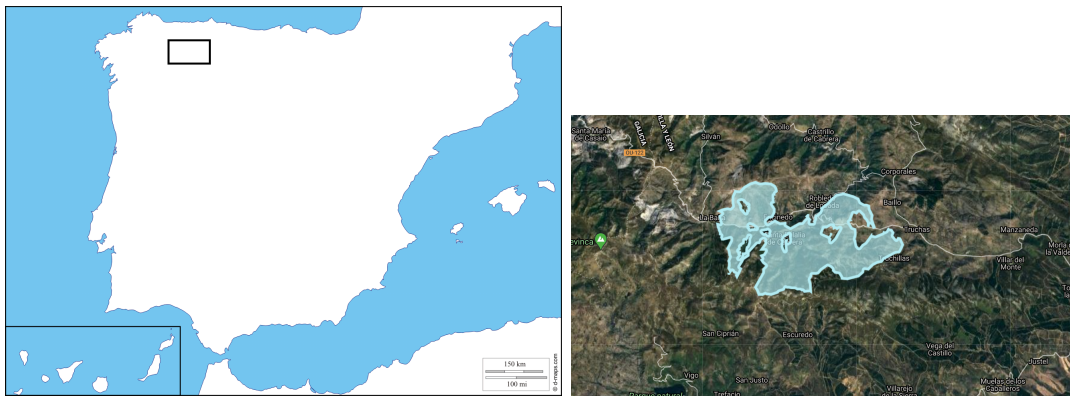


Figura 1: Perímetro incendio de Encinedo.

Fuente: <http://almazcara.forestry.es/2017/08/incendio-en-la-cabrera.html> .

### 1.3.2. Galicia

Galicia es una comunidad autónoma situada en el noroeste de la península ibérica. El clima de Galicia es, principalmente, oceánico. Este varía según la zona donde nos encontremos entre suboceánico y mediterráneo. Dentro de la flora destaca su gran porcentaje de monte (69,96 %) y bosque. Con especies que varían según el clima, donde se pueden destacar los bosques de robles, las estepas y los tejos entre otros. Los incendios de Galicia de octubre de 2017 forman parte de una serie de incendios que afectaron además de a Galicia, a Asturias, Castilla y León y el norte y centro de Portugal. Fueron arrasadas unas 50000 *ha* según la junta de Galicia [16].

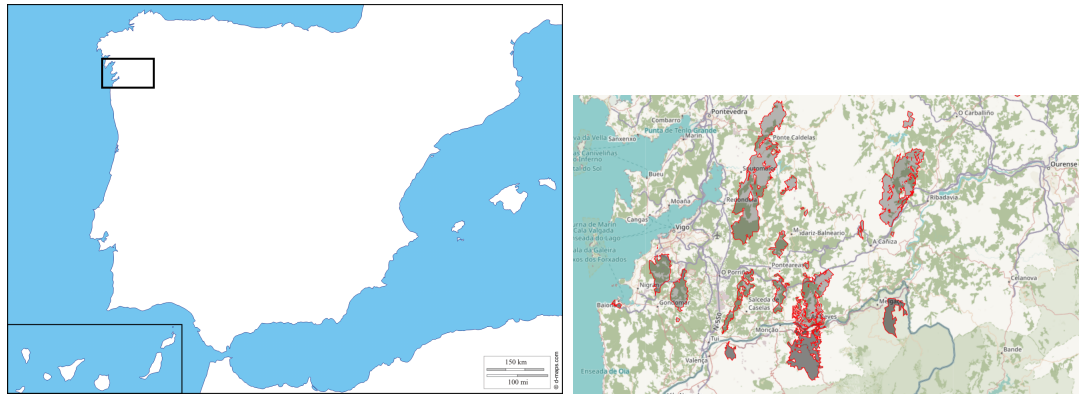


Figura 2: Principales incendios de Galicia en octubre de 2017.

Fuente: <http://almazcara.forestry.es/2017/10/incendios-en-espana-y-portugal.html>.

## 2. Teledetección Satelital

### 2.1. Teledetección

Se puede definir la teledetección como la adquisición de información sobre el estado y condición de un objeto a través de sensores que no están en contacto físico con él [18]. En este proceso entran en juego interacciones de energía entre el objetivo y el sensor. La señal que detecta el sensor puede ser energía electromagnética proveniente del Sol que es reflejada por la superficie terrestre o bien energía emitida por la propia Tierra. También se pueden fabricar sensores que produzcan su propia fuente de energía (por ejemplo, un radar, que utiliza como fuente de energía ondas de radio), pudiendo captar la señal tras interactuar esta con la superficie. Para todos los casos, la señal de energía es medida por el sensor y posteriormente guardada en la memoria a bordo del satélite o transmitida a una estación terrestre para su interpretación. Y esto es importante, la teledetección también abarca el análisis y la interpretación de los datos e imágenes adquiridos.

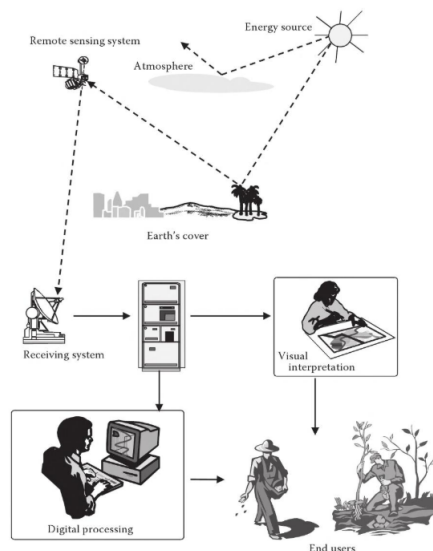


Figura 3: Esquema resumen del proceso de teledetección explicado previamente [17].

En resumen, la teledetección incluye seis componentes:

1. Una fuente de energía o de radiación electromagnética que va a ser detectada por un sensor. En nuestro caso esa fuente de energía es el Sol.
2. La superficie terrestre, formada por vegetación, agua rocas, nieves, suelos y estructuras creadas por el ser humano. La superficie recibe la energía solar y, como resultado de interacciones energía-materia, refleja y emite parte de esa energía hacia el sensor del satélite.
3. Los sensores y tecnología para el proceso de la teledetección.

4. La estación que recibe los datos obtenidos, los cuales han de ser procesados.
5. Los y las analistas que transforman los datos procesados previamente en información interpretable y de interés.
6. Y por último, la comunidad de usuarios que utilizan los resultados para diferentes aplicaciones.

### **2.1.1. El programa Copernicus**

El programa Copernicus es una iniciativa liderada por la Comisión Europea en colaboración con la Agencia Espacial Europea ESA. Este programa tiene como misión construir un sistema autónomo de observación de la Tierra en base a: la observación espacial utilizando una red de satélites, la observación in-situ mediante redes de estaciones de medida en tierra y medios aéreos, y la generación de servicios de información. El objetivo general es proveer de información exacta, fiable y continua, para mejorar la gestión y conservación del medio ambiente, comprender y mitigar los efectos del cambio climático. La información que proporciona Copernicus se agrupa en cinco categorías: la Tierra, la atmósfera, los océanos, el cambio climático y respuestas a emergencias y seguridad. El programa ofrece datos y servicios de información de forma gratuita y abierta [19].

El programa Copernicus utiliza una familia de satélites denominados Sentinel. El primer satélite de esta familia es el Sentinel-1A y fue lanzado en 2014. También se apoya en sistemas de medición in-situ que proporcionan datos útiles para calibrar, verificar y complementar la información obtenida por los satélites. Dentro de la familia de satélites, el Sentinel-2 está destinado a proveer imágenes ópticas de alta resolución para servicios terrestres. Sentinel-2 es la misión utilizada en este trabajo.

Sentinel-2 es una misión formada por dos satélites, el Sentinel-2A y el Sentinel-2B que tiene como objetivo la monitorización de cambios en la superficie terrestre. Para ello, obtiene imágenes multiespectrales y multitemporales de un determinado área de la superficie [20]. Ambos satélites tienen las mismas características técnicas y comparten misma órbita polar pero con una diferencia de fase de  $180^\circ$ , lo que permite tener un tiempo entre dos observaciones del mismo punto terrestre de 5 días. Para la obtención de información, los satélites portan un sensor (MSI) que muestrea un total de 13 bandas espectrales, divididas en tres grupos según su resolución de 10, 20 o 60 metros. El sensor proporciona cobertura global de la superficie terrestre y de regiones costeras, cubriendo las regiones espectrales propias al visible, cercanas al infrarrojo y de onda corta del infrarrojo. Cada banda mide en un intervalo del espectro electromagnético aportando su propia información. Para visualizar y clarificar cuáles y cómo se disponen estas bandas se presenta un cuadro resumen de ellas en la Tabla 1:

Sentinel2 Bands	Central Wavelength ( $\mu m$ )	Resolución ( $m$ )
Band 1: Coastal aerosol	0.443	60
Band 2: Blue	0.490	10
Band 3: Green	0.560	10
Band 4: Red	0.665	10
Band 5: Vegetation Red Edge	0.705	20
Band 6: Vegetation Red Edge	0.740	20
Band 7: Vegetation Red Edge	0.783	20
Band 8: NIR	0.842	10
Band 8A: Vegetation Red Edge	0.865	20
Band 9: Water vapour	0.945	60
Band 10: SWIR - Cirrus	1.375	60
Band 11: SWIR	1.610	20
Band 12: SWIR	2.190	20

Tabla 1: Tabla resumen de las bandas de Sentinel-2 junto su longitud de onda central y su resolución [21].

### 2.1.2. Principios físicos de la teledetección

En teledetección, la información sobre el estado y las características de un objeto se extrae interpretando la señal electromagnética proveniente del objeto. La radiación solar reflejada por la superficie terrestre puede servirnos para conocer las propiedades físicas, químicas y biológicas del suelo, hielo, nieve, agua y vegetación de un área. El objetivo de la teledetección es entender como la energía electromagnética interacciona con la superficie para poder extraer información relevante.

Dentro de la teledetección existen tres formas de recoger información sobre un objeto: reflexión, emisión y la combinación de emisión-reflexión. Entre ellas la reflexión es la más importante porque utiliza luz solar, que es la fuente de energía principal de la Tierra. El Sol ilumina la superficie terrestre y esta refleja una parte de esa energía de vuelta al espacio dependiendo del tipo y de la composición de la cobertura presente en la superficie. La porción reflejada es captada por el sensor a bordo del satélite en forma de datos que son transformados en información. Esta información proviene en forma de radiación electromagnética y por lo tanto es necesario entender las propiedades de este tipo de radiación para saber el por qué de los resultados.

Las propiedades de la radiación electromagnética pueden ser explicadas a través de dos teorías: la teoría de ondas y la teoría cuántica. Según la teoría de ondas, la radiación electromagnética es el resultado de la oscilación de los campos magnéticos y electrostáticos, los cuales son ortogonales entre sí y a la dirección de propagación [22]. La energía electromagnética es transmitida de un lugar a otro siguiendo un modelo continuo con una velocidad igual a la velocidad de la luz,  $c$ . Las propiedades de la energía se pueden entender entonces mediante la ecuación que define esta velocidad:

$$c = \lambda \nu \quad (1)$$

donde  $c$  es la velocidad de la luz,  $\lambda$  la longitud de onda y  $\nu$  la frecuencia.

Por otro lado, la teoría cuántica de la luz describe la radiación como una sucesión de paquetes de energía denominados fotones, cuya masa es nula [23]. La cantidad de energía que puede transportar un fotón en relación a su frecuencia es:

$$E = h\nu \quad (2)$$

donde  $E$  es la energía radiante de un fotón,  $\nu$  la frecuencia y  $h$  la denominada constante de Planck.

Si combinamos las dos últimas ecuaciones tenemos que:

$$E = \frac{hc}{\lambda} \quad (3)$$

lo que determina que cuanto mayor es la longitud de onda menor es la energía contenida y vice versa. Esto supone que la radiación de longitud de onda grande es más difícil de detectar que la radiación de longitud de onda corta. Llegados a este punto, conviene hablar más específicamente sobre el espectro electromagnético.

### 2.1.3. El espectro electromagnético

Como hemos visto la radiación de energía se puede caracterizar en función de su longitud de onda o frecuencia. El espectro electromagnético muestra el conjunto de longitudes de onda y frecuencias y las agrupa en distintas regiones. Dentro de estas regiones, son tres las más utilizadas en teledetección [24]:

1. La región visible (VIS), de 0.4 a 0.7  $\mu m$ . Esta es básicamente la parte de la radiación electromagnética que nuestros ojos son capaces de detectar.
2. La región cercana al infrarrojo (NIR), de 0.7 a 1.2  $\mu m$ . El interés principal de esta región reside en su capacidad de discriminar el verde de la vegetación.
3. La región media-infrarrojo (MIR), de 1.2 a 8  $\mu m$ . En concreto dentro del intervalo que va desde 1.2 a 3  $\mu m$  concierne la radiación solar reflejada por la superficie terrestre, esta región se denomina SWIR. El resto de la región MIR es útil para estimar la humedad de suelos y vegetación así como detectar fuentes de altas temperaturas.

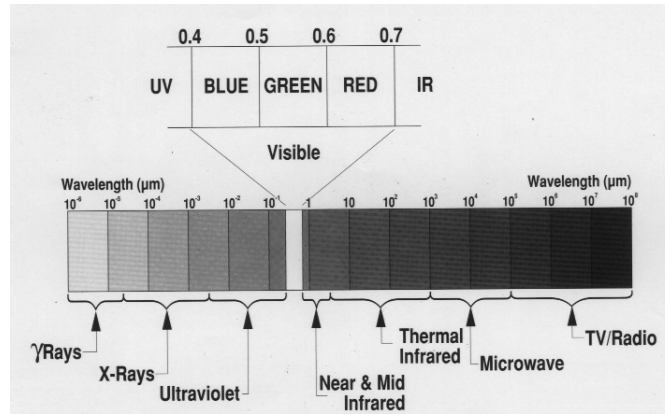


Figura 4: Espectro electromagnético. Fuente: <https://casde.unl.edu/tutorial/rs-intro/spectrum/index.php>.

#### 2.1.4. Firma espectral

Las señales de energía electromagnética recibidas por el sensor satelital pueden ser de distintas regiones del espectro electromagnético. Esto vendrá determinado por las propiedades y el comportamiento del objeto sobre las distintas longitudes de onda, lo que se conoce como firma espectral [25]. En la Figura 5 se muestran las firmas espectrales de varios elementos que podemos encontrarnos en la superficie terrestre. Se puede observar la diferencia en el comportamiento espectral según el elemento.

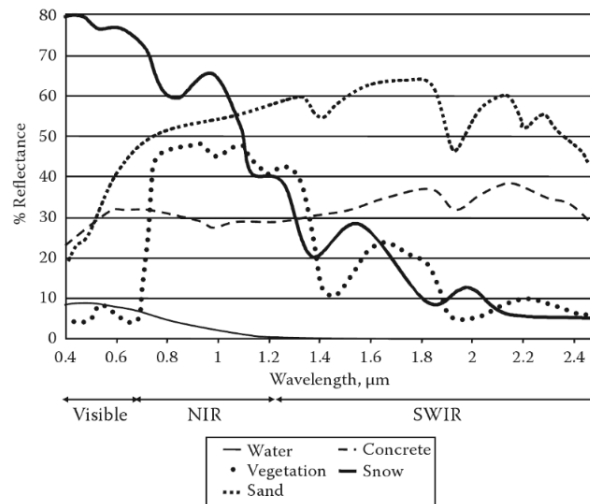


Figura 5: Firma espectral para distintos materiales [26].

Es importante tener en cuenta que existen factores medioambientales que pueden influenciar la energía recibida en el sensor y variar los datos afectando a su interpretación. Condiciones como el estado de la atmósfera, la localidad y topografía del área o la calibración de los instrumentos con los que vamos a obtener

los datos pueden generar ruido en las medidas. Si bien es cierto que, a lo largo de los años la tecnología ha avanzado y se han mejorado los dispositivos a bordo de los satélites, permitiendo utilizar entre otras cosas, algoritmos que corrijan posibles variaciones en las medidas.

En definitiva, la firma de reflectancia espectral es muy útil para extraer información sobre la superficie terrestre y analizar imágenes. La firma espectral de la vegetación es, en concreto, la que más interesa de cara al análisis de este trabajo. Es la base de los índices que son utilizados para la obtención de resultados y es por ello que toca hablar sobre ello.

## 2.2. Índices de Vegetación

Si se observa la Figura 5 se puede ver que la firma espectral de la vegetación tiene poca reflectancia tanto en la región visible como en la región SWIR del espectro electromagnético, mientras que tiene alta reflectancia en la región NIR. Este comportamiento depende de varios y complejos factores. Principalmente las propiedades de reflectancia de la cubierta vegetal dependen de los niveles de pigmentos fotosintéticos, contenido en agua y estructura [27]. La baja reflectancia en la región VIS se debe a la absorción de la clorofila. La absorción se produce en la zona del visible, en concreto en la región del azul y del rojo. La zona del visible propia al verde no absorbe la radiación sino que la refleja, por eso el color verde de la vegetación. En la región NIR la reflectancia es mayor que en la VIS debido a la estructura celular de las hojas que refleja la mayor radiación del NIR. El descenso final de la reflectancia propio de la región SWIR está relacionado con el contenido de agua de la vegetación ya que el agua absorbe fuertemente la radiación entre 1.45 y 2.50  $\mu m$ .

La información satelital utilizada para analizar la cubierta vegetal de un área comúnmente viene dada en una imagen multibanda. Con multibanda nos referimos al uso de distintas bandas espectrales de los satélites, donde cada banda capta una zona del espectro electromagnético. La imagen ha de ser categorizada y agrupada para poder discriminar un área con características distintas a otra área. Una forma de realizar este agrupamiento es utilizando índices.

Podemos entender los índices como el conjunto de operaciones matemáticas sobre los valores de los píxeles de las imágenes. En concreto, un índice de vegetación es el resultado de las operaciones matemáticas utilizando los valores de la reflectancia a distintas longitudes de onda sobre la cubierta vegetal. Existen numerosos índices de vegetación, que sirven para analizar distintas características de la superficie. Entre ellos, el índice NDVI y NBR son los principales para estudiar la cubierta afectada por un incendio forestal. Estos son los utilizados en este trabajo.

### 2.2.1. Índice NDVI

El Índice de Vegetación Diferencial Normalizado es el más utilizado para todo tipo de situaciones en la agricultura, topografía, para mapear superficies y en general para análisis de vegetación [28]. Este varía entre -1 y 1 donde 0 es el punto en



el cual empieza a haber ausencia de vegetación. Los valores negativos representan zonas sin vegetación. La normalización reduce el efecto de degradación en la calibración del sensor y disminuye los efectos atmosféricos. Para su cálculo se utilizan la banda cercana al infrarrojo (NIR) y la banda que mida en el rojo (RED). Son estas las utilizadas debido a que las longitudes de onda cercanas al infrarrojo (NIR) son fuertemente reflejadas por la vegetación mientras que las longitudes de onda dentro del rojo (RED) son fácilmente absorbidas. El índice NDVI es definido de la siguiente forma:

$$NDVI = \frac{RED - NIR}{RED + NIR} \quad (4)$$

Si la vegetación es sana (alto contenido de clorofila) la reflectancia en NIR es alta y en RED baja por lo que se tienen valores de NDVI positivos y cercanos al 1. Pasa lo contrario si la vegetación no es sana, teniendo valores próximos al 0. En la Tabla se presentan los valores del índice NDVI utilizados para identificar el tipo de vegetación de la zona estudiada.

Tipo de cobertura	NDVI rango
Nubes, nieve, agua , zonas áridas	(-1.0,0.0)
Vegetación poco densal	(0,0.3)
Vegetación densa	(0.3,0.6)
Vegetación muy densa	(0.6,1.0)

Tabla 2: Relación de tipo de vegetación o terreno y valores del índice NDVI [29], [30].

### 2.2.2. Índice NBR

El índice de calcinación normalizada (NBR) sirve para analizar la severidad de un incendio en una región. El comportamiento de las masas vegetales en función del espectro electromagnético dentro del infrarrojo es lo que permite hacer los estudios forestales pertinentes a los incendios. Como se ha mencionado en el anterior apartado referente al índice NDVI, la vegetación sana tiene alta reflectividad en la banda NIR y va disminuyendo cuando nos alejamos de esta banda hacia la banda RED o a la banda SWIR (luz infrarroja de onda corta). Lo contrario ocurre para la vegetación afectada en un incendio; poca reflectividad en la banda NIR y alta en la banda SWIR. De esta forma se define el índice NBR como [31]:

$$NBR = \frac{NIR - SWIR}{NIR + SWIR} \quad (5)$$

Al igual que el índice NDVI, el NBR tiene valores que van desde el -1 al 1. Los valores negativos representan las zonas afectadas por el incendio mientras que los positivos representan las zonas sanas o en fase de regeneración. Además, mediante la comparación de dos o más índices NBR de la misma zona en distinta fecha,

se puede determinar el grado de severidad del incendio, lo que se denomina como índice dNBR.

El índice dNBR o difference Normalized Burn Ratio es la diferencia del índice NBR pre- y post-incendio [32].

$$dNBR = NBR_{post} - NBR_{pre} \quad (6)$$

Su cálculo permite hacer una evaluación inicial de la severidad del fuego, si se utiliza una imagen realizada cuando el incendio se ha extinguido, o una evaluación amplia, si la imagen es posterior al inicio de la regeneración de vegetación de la zona afectada. Al igual que los índices NDVI y NBR su intervalo de valores va desde -1 hasta 1. Los valores negativos representan zonas de regeneración de la vegetación y los positivos representan zonas más afectadas. Si bien es cierto que el dNBR puede variar según el caso estudiado y su interpretación se tiene que adaptar a este. Se recoge, en la Tabla 3, los intervalos de valores utilizados para analizar este índice.

Gravedad	dNBR
Recrecimiento, alto	(-0.50,-0.251)
Recrecimiento, bajo	(-0.25,-0.101)
Sin quemar	(-0.10,0.10)
Baja severidad	(0.10,0.27)
Media severidad	(0.27,0.44)
Alta severidad	(0.44,0.67)

Tabla 3: Nivel de severidad y su valor para el índice dNBR [33].

### 3. Metodología y Técnicas Experimentales

#### 3.1. Adquisición y procesamiento de datos

Los datos utilizados provienen del Copernicus Open Access Hub, portal que proporciona de forma totalmente libre y abierta datos completos de las misiones Sentinel-1, Sentinel-2, Sentinel-3 y Sentinel-5P. En este trabajo los datos son de la misión Sentinel-2. Alguien que quiera extraer datos de cualquiera de las misiones mencionadas lo único que tiene que hacer es registrarse vía online en el portal. Hay que seleccionar el tipo de producto (nivel 2A), las coordenadas y fechas deseadas.

Los datos adquiridos vienen en forma de matriz o, lo que es igual, de imagen. Estos elementos tienen un determinado valor numérico en función del cual el píxel se representa de un color. Dicho valor numérico no es más que la reflectividad medida por el sensor del Sentinel-2. En resumen, el sensor del satélite capta la reflectividad de un área y a través de ella forma la imagen de ese área.

Los datos de Sentinel-2 disponibles a usuarios están divididos en dos niveles: nivel 1C y nivel 2A. El nivel 2A proporciona valores de reflectancia en la parte baja de la atmósfera. Se aplican correcciones atmosféricas al nivel 1C. Son los datos de este nivel 2A los utilizados en este trabajo.

Las imágenes del Nivel-2A se disponen en baldosas de  $100 \times 100 km^2$  según la proyección UTM, cada una de las cuales pesa aproximadamente 500 MB. La proyección UTM (Universal Transverse Mercator) divide la superficie terrestre en 60 zonas. Cada una tiene una vertical de  $6^\circ$  de longitud y una horizontal de  $8^\circ$  de latitud. Las imágenes están comprimidas en formato JPEG2000, todas ellas encapsuladas de nuevo dentro de un contenedor de archivos SAFE.

#### 3.2. Métodos empleados para la detección de los incendios

En los modelos utilizados para la detección de incendios, el objetivo es obtener una máscara de la zona del incendio. Por máscara nos referimos a una imagen donde solo se represente la zona incendiada. Para crear estas máscaras se utiliza una imagen satelital del área de interés, se calculan los índices NBR y NDVI mediante la combinación de bandas espectrales pertinentes y, en función de los valores bibliográficos, se representan solo los píxeles que estén por debajo del valor umbral.

Una vez obtenidas las máscaras se cuenta el número de píxeles incendio. Estos van a ser utilizados para verificar la eficacia de los modelos utilizados. La forma de contar los píxeles es la siguiente: la máscara se forma por los píxeles incendio con el valor 1 y el resto con el valor 0. Se recorre la imagen píxel por píxel, contando aquellos con valor 1. De esta forma, se puede calcular el ratio  $r$  entre los píxeles

incendio  $n$  de los modelos utilizados y los píxeles incendio de las máscaras originales  $N$ :

$$r = \frac{n}{N} \quad (7)$$

### 3.2.1. Resta de imágenes

Este método consiste en la resta, píxel por píxel, de dos imágenes, dando como resultado una nueva imagen que representa los cambios existentes entre ambas. Aplicado a nuestro caso, se utiliza una imagen previa al incendio y otra posterior, tanto para el NBR como para el NDVI. La diferencia entre la imagen posterior y la anterior debería mostrar la zona del incendio, pues es el área que más ha cambiado entre las dos fechas. Con dicha zona se crea la máscara incendio para cada índice.

De los tres métodos realizados en este trabajo es el más directo. Ahora bien, tiene sus desventajas. Principalmente la alta sensibilidad a errores en el registro geométrico y a la existencia de píxeles mezcla. Además se necesita la correcta definición de un umbral que marque las diferencias entre las imágenes. Para reducir estos errores es necesario tratar los datos con anterioridad. La aplicación de máscaras para eliminar nubes y agua y la elección correcta del umbral que dictamine la zona de incendio y no incendio son de suma importancia para obtener resultados consistentes.

Un ejemplo sencillo de como funciona este método es el propio a la Figura 6 :



Figura 6: Ejemplo del método de resta de imágenes. Fuente: <https://www.pyimagesearch.com/2017/06/19/image-difference-with-opencv-and-python/>.

Las dos imágenes de la parte superior son las que se restan. La diferencia entre ambas es que una tiene el logo de la marca de la tarjeta mientras que la otra no. Las dos imágenes de la parte inferior muestran el resultado de la resta, apreciándose el logo que se quería detectar.

### 3.2.2. Aprendizaje no supervisado: K-Means

K-Means es un algoritmo de clasificación no supervisada (clusterización) que localiza y agrupa en clases o clusters los puntos de datos que tienen similitud entre ellos [34]. En el algoritmo, similitud es lo mismo que cercanía en la distancia entre puntos. Cuanto más cerca estén los puntos mayor similitud entre ellos y mayor probabilidad de pertenecer al mismo clúster. El agrupamiento se efectúa minimizando la suma de distancias entre cada punto y el centroide de su cluster. Los puntos se representan con vectores reales de  $d$  dimensiones  $(x_1, x_2, \dots, x_n)$  y el algoritmo crea  $k$  clusters donde se minimiza la suma de las distancias de los puntos a su centroide dentro de cada grupo  $S = (S_1, S_2, \dots, S_k)$ .

El objetivo del K-Means es minimizar el factor de inercia del cluster, para ello se realizan los siguientes pasos:

1. Elegimos el número de clústers  $k$  que queremos.
2. El algoritmo elige aleatoriamente los centroides de cada grupo.
3. Se asigna cada punto de datos al centroide más cercano.
4. Se calcula la inercia de los grupos.
5. Se calculan nuevos centroides a través de la media de los puntos que pertenecen al centroide del anterior paso.
6. Se repite el proceso desde el paso 3 hasta que los centroides no se desplacen o si lo hacen por debajo de una distancia umbral.

El planteamiento matemático es el siguiente:

$$\min E(\mu_i) = \min \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (8)$$

donde  $S$  es el conjunto de datos cuyos elementos son los puntos  $x_j$  representados por vectores. Tendremos  $k$  clusters con  $\mu_i$  como centroides de dichos clusters.

En cada cambio de los centroides se impone la condición de extremo a la función  $E(\mu_i)$ , lo cual da como resultado:

$$\frac{\partial E}{\partial \mu_i} = 0 \quad (9)$$

$$\mu_i = \frac{1}{|S_i|} \sum_{x_j \in S_i} x_j \quad (10)$$

y se toma el promedio de los elementos de cada grupo como nuevo centroide.

Un aspecto de suma relevancia en K-Means es la elección del número de clusters  $k$ . Una forma de hacer esto es el elbow method. Este se utiliza para determinar el número correcto de grupos en un dataset. Para ello se representan los valores ascendentes de  $k$  frente a la varianza obtenida con dicho  $k$ . El objetivo es encontrar el  $k$  óptimo, este es el menor posible para una varianza pequeña. La varianza se define como:

$$\%Varianza = \frac{VarianzaGrupos}{VarianzaTotal} \quad (11)$$

Un ejemplo aplicado es el siguiente:

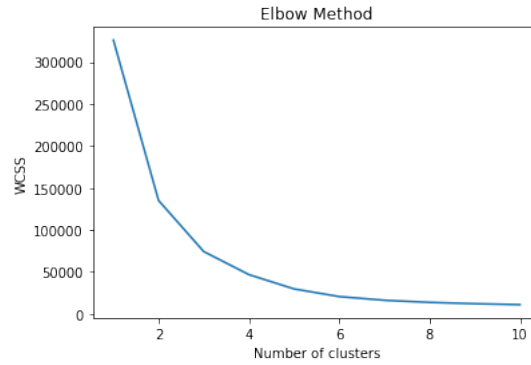


Figura 7: Ejemplo del método del codo utilizando los datos de los índices NBR y NDVI del incendio de Enciendo.

En este caso se elegiría  $k = 5$  ya que a partir de este número de clusters la varianza apenas cambia, el resto de clusters por encima de  $k = 5$  mantienen aproximadamente el mismo valor.

Hay que tener en cuenta que si bien K-Means es un algoritmo de agrupación efectivo, es más útil cuando sabemos el número concreto de grupos que necesitamos y cuando las distribuciones de datos son esféricas.

### 3.2.3. Aprendizaje supervisado: Regresión logística

El algoritmo de regresión logística es uno de los más utilizados en aprendizaje supervisado [35]. Tiene como principal aplicación los problemas de clasificación binaria, estos son aquellos en los que existen dos posibles resultados (sí o no, 1 o 0). A pesar de denominarse como regresión no es aplicable en problemas de esta índole, en los cuales se busca un valor continuo, sino que es una técnica para problemas de clasificación.

Con la regresión logística se mide la relación entre la variable dependiente, lo que se desea predecir, con una o más variables independientes, las características para el modelo. Para hacer esto se utiliza una función logística que determina la

probabilidad de la variable dependiente, probabilidad que es traducida en valores binarios. La función utilizada es la función sigmoide. Esta es una curva que puede tomar únicamente valores entre 0 y 1. Su ecuación es la siguiente:

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (12)$$

donde  $x$  es un número real.

En esta ecuación, cuando  $x$  tiende a menos infinito  $f(x)$  tiende a cero, mientras que cuando  $x$  tiende a infinito  $f(x)$  tiende a la unidad. En la Figura 8 se representa gráficamente el comportamiento de la función sigmoide.

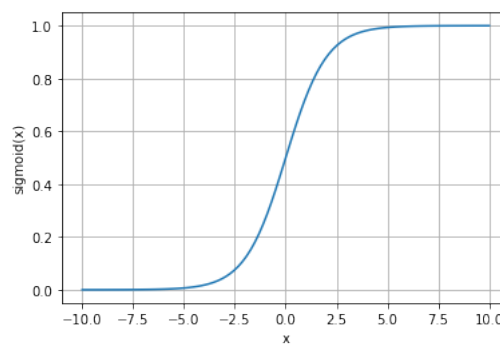


Figura 8: Función sigmoide. Creación propia usando Python.

La regresión logística tiene como principal ventaja su alta interpretabilidad de los resultados. El peso de cada característica determina los resultados finales, por lo que se puede afirmar que el modelo toma una u otra decisión en función del uso de sus características. Como desventaja está la imposibilidad de resolver problemas no lineales, lo cual se debe a que la expresión que toma la decisión es lineal. Además, la variable objetivo tiene que ser linealmente separable, en los datos tiene que haber dos regiones con una frontera lineal.

### 3.2.4. Parámetros de validación: la matriz de confusión

La matriz de confusión es una tabla que permite visualizar y analizar la actuación de un algoritmo [36]. Para entender cómo funciona lo mejor es representarla:

		Actual class	
		P	N
Predicted class	P	TP	FP
	N	FN	TN

Figura 9: Ejemplo de una matriz de confusión. Siendo TP: verdaderos positivos, TN: verdaderos negativos, FP: falsos positivos y FN: falsos negativos.

La matriz se compone de dos filas que representan las predicciones de cada clase y dos columnas que representan las propias clases. Los valores verdaderos son los satisfactoriamente predichos por el método que se haya utilizado mientras que los valores falsos son los predichos equívocamente [37]. De esta forma es fácil saber cómo actúa nuestro método si utilizamos parámetros estadísticos que proporcionen información de los resultados. Los empleados en este trabajo son el índice de Matthews MCC, la sensibilidad  $S$  y la precisión  $P$ .

El coeficiente de correlación de Matthews MCC, es utilizado como medida de la calidad de clasificaciones binarias y multiclases [38]. Tiene en cuenta los verdaderos y falsos positivos y negativos. El MCC tiene valores comprendidos entre -1 y 1, siendo -1 una predicción inversa, 0 una típica predicción aleatoria y 1 una predicción perfecta. La idea del MCC es que cuanto mayor sea la correlación entre los valores verdaderos y los predichos, mejor predicción se ha hecho. Se define matemáticamente de la siguiente forma:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (13)$$

Si se observa la ecuación, cuando  $FP=FN=0$  el valor de MCC es 1, esto es cuando la clasificación es perfecta ya que no existen valores falsos. En cambio, cuando  $TP=TN=0$  el resultado de MCC es -1, no existen valores verdaderos. Una cualidad importante del MCC es su simetría, una clase no es más importante que la otra, se consideran los valores positivos y los negativos por igual.

La sensibilidad  $S$ , es el número de verdaderos positivos dividido por el número total de positivos. Es la proporción de positivos correctamente identificada como tal:

$$S = \frac{TP}{TP + FN} \quad (14)$$

La precisión  $P$ , es el número de verdaderos positivos dividido por el número total de predicciones positivas. La precisión es útil para determinar si el coste de Falsos Positivos es alto:



$$P = \frac{TP}{TP + FP} \quad (15)$$

### 3.3. Procedimiento Empleado

Antes de haber comenzado todo el proceso de obtención de resultados ha sido necesario un periodo de familiarización y aprendizaje del lenguaje de programación utilizado, Python. En concreto de librerías destinadas al análisis de datos como Numpy, Pandas, SciKit y Matplotlib. También el manejo de datos NetCDF y como operar con los arrays de las bandas satelitales.

#### 3.3.1. Detección de los incendios

La detección de los incendios es la parte principal del trabajo y la que más labor supone. Lo primero es descargar los ficheros SAFE de Sentinel-2 de cada zona del incendio. Para descargar estos archivos se utiliza un programa (consultar glosario) desarrollado por el equipo de computación del IFCA que permite descargar y transformar los archivos a formato NetCDF. En el programa el usuario tiene que introducir las coordenadas y las fechas deseadas. Los archivos a descargar son de una fecha previa al incendio y de otra justo posterior a la extinción de este. Una vez se tienen los archivos se escribe el código donde se aplican los tres modelos empleados para la detección de incendios. El código realizado actúa de la siguiente forma:

1. Se cargan las bandas que van a ser utilizadas para el cálculo de los índices. Las bandas son de las resoluciones 10 m y 20 m. La resolución 60 m no es necesaria en nuestro caso.
2. Se realiza una transformación en las dimensiones de las bandas de resolución 10 m para que tengan la misma dimensión que las bandas de resolución 20 m. Esto permite operar entre bandas de resolución diferente. Además, en el caso de León, se ajustan los arrays a la zona del incendio. En Galicia no es necesario porque se utiliza toda la imagen.
3. Se calculan primero los índices NSI y NDWI para obtener las máscaras de agua y nubes respectivamente. Se calculan los índices NBR y NDVI y se aplican las máscaras anteriores. De esta forma en cada imagen se desprecian tanto nubes como zonas de agua que puedan entorpecer la interpretación de los resultados.
4. Con las imágenes obtenidas tras el paso 3 se consiguen las denominadas máscaras originales. Para ello se consideran todos los valores por encima de un valor umbral como incendio (valor igual a 1) y el resto como no incendio (valor igual a 0). Para el índice NBR el valor umbral estimado es  $-0,2$  y para el índice NDVI es 0. Las máscaras originales son utilizadas para comparar los resultados de los tres modelos.

Tras este proceso se aplican los modelos de resta de imágenes, K-Means y regresión logística.

La resta de imágenes utiliza los archivos previo y posterior al incendio. Siguiendo los pasos descritos anteriormente se consiguen las imágenes NBR y NDVI para cada caso. Se realiza la resta entre la imagen posterior y la imagen anterior para cada índice. Como resultado se tiene la zona del incendio pues es la que ha sufrido más cambios en el tiempo.

En el método no supervisado, K-Means, se ha utilizado en un mismo vector los datos NBR y NDVI. A diferencia de la resta de imágenes y del método supervisado, en los que se hace su procedimiento de forma independiente para cada índice. Es decir, como resultado se tiene una máscara. Los pasos que sigue el código de este modelo están explicados en la sección 3.2.2. Para saber qué número de clústers,  $k$ , es el que mejor resultados genera, se prueba con distintos valores y se analizan los resultados pertinentes. En el capítulo de Resultados se encuentran los detalles. Además, este modelo requiere que el usuario separe el clúster incendio para crear la futura máscara. Para ello, se examinan los valores de los centroides. El centroide cuyo valor sea el más próximo a  $(-0,2)$ , valor umbral del NBR, es el centroide del clúster incendio, que es utilizado para crear la máscara final.

El método supervisado, regresión logística, utiliza por un lado los datos del índice NBR y por otro los del índice NDVI. Se consigue una máscara para cada uno. Uno de los parámetros de entrada del algoritmo es la proporción de datos de entrenamiento y de testeo. De la misma forma que con el método no supervisado, se prueba con distintas proporciones y se comprueba cuál es la mejor. Como resultado se obtiene directamente la máscara incendio predicha por el modelo.

Para cada método utilizado se emplea el ratio  $r$  y la matriz de confusión y sus parámetros, el índice MCC, la sensibilidad  $S$  y la precisión  $P$ , para analizar los resultados. Para calcular la matriz de confusión es necesario comparar la máscara original con la máscara obtenida por cada modelo. Las máscaras son un array de ceros y unos. Cero es zona no incendiada y uno es zona incendiada. Los unos de la máscara predicha que también son unos en la original son verdaderos positivos. Lo mismo pero para ceros son verdaderos negativos. Los unos en la máscara predicha que son ceros en la original son falsos positivos y al revés son falsos negativos. De la matriz de confusión se utilizan los verdaderos positivos como resultado para futuros cálculos como el ratio entre máscaras y las hectáreas. Una vez hallada la matriz de confusión se determina el índice MCC mediante la Ec 13, la sensibilidad  $S$  mediante la Ec 14 y la precisión  $P$  mediante la Ec 15.

### 3.3.2. Avance temporal de los incendios

El propósito principal de esta parte es visualizar cómo evolucionan los índices utilizados en el trabajo desde que se extinguen los incendios. Además se pretende comparar las evoluciones de Encinedo y Galicia desde el punto de vista de la vegetación. Para ello, primero se elige el modelo que mejor resultados genere de

la sección de detección de incendios. Asegurando así que los resultados para esta sección sean lo más precisos posibles. Los pasos a seguir son los siguientes:

1. Descargar archivos de Sentinel-2 de la zona del incendio, desde su origen hasta el presente. Intentando conseguir un archivo por mes. Los archivos que contengan exceso de ruido en la imagen son descartados.

2. Aplicar el modelo de regresión logística para cada fecha, obteniendo la máscara incendio. El modelo es previamente entrenado y validado con los datos de la sección de detección de incendios, no es necesario entrenar el modelo para cada archivo.

3. Obtener las imágenes NBR y NDVI para las zonas de regeneración. Para ello se calculan los índices y se aplican la máscara original de cada uno y la máscara incendio predicha mediante el método supervisado.

4. Obtener el número de píxeles para cada intervalo del índice dNBR y NDVI, lo que permite realizar un análisis de la evolución temporal del incendio de forma numérica. Es importante destacar que los valores empleados son del índice dNBR y no del NBR. Para obtener el índice dNBR se resta la imagen NBR de la fecha empleada y la imagen previa al incendio siguiendo la Ec 6.

5. Por último, comparar la evolución del incendio de Encinedo y uno de los incendios de Galicia en función de la vegetación de cada zona.

Para este último punto se ha utilizado el portal de datos de biodiversidad GBIF (Global Biodiversity Information Facility), el cual permite visualizar estadísticas de la vegetación de un área.

## 4. Resultados y Análisis

Los códigos de Python con los que se han obtenido los resultados, tanto de detección de incendios como de la evolución temporal de la superficie afectada por el incendio, están disponibles en [39].

### 4.1. Detección de los incendios

#### 4.1.1. Máscaras Originales: Encinedo

Para comparar los resultados de los métodos utilizados en esta sección se utilizan las máscaras originales, una con valores NBR y otra con valores NDVI. El procedimiento para obtenerlas viene explicado en la sección anterior. Las máscaras originales del incendio de Encinedo aparecen en la Figura 10 .

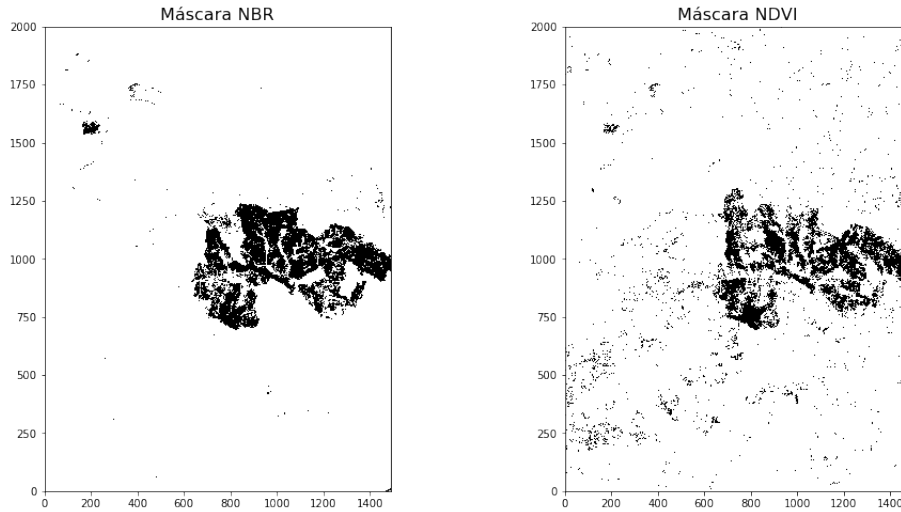


Figura 10: Máscaras originales que van a servir para comparar los resultados de los tres modelos usados para la detección del área incendiada de Encinedo.

Los píxeles identificados como incendio en cada máscara:

$N_{nbr}$	$N_{ndvi}$
192428	196898

Tabla 4: Número de píxeles incendio de la máscara original NBR y de la máscara original NDVI de Encinedo.

La máscara incendio con el NBR es más limpia que la máscara con el NDVI ya que esta última presenta píxeles sueltos, identificados como incendio, repartidos

por toda la imagen. Estos píxeles sueltos de la máscara NDVI se traducen en mayor número de píxeles identificados como incendio en comparación con el NBR.

#### 4.1.2. Máscaras Originales: Galicia

Las máscaras originales referentes a los incendios de Galicia, se presentan en la Figura 11 .

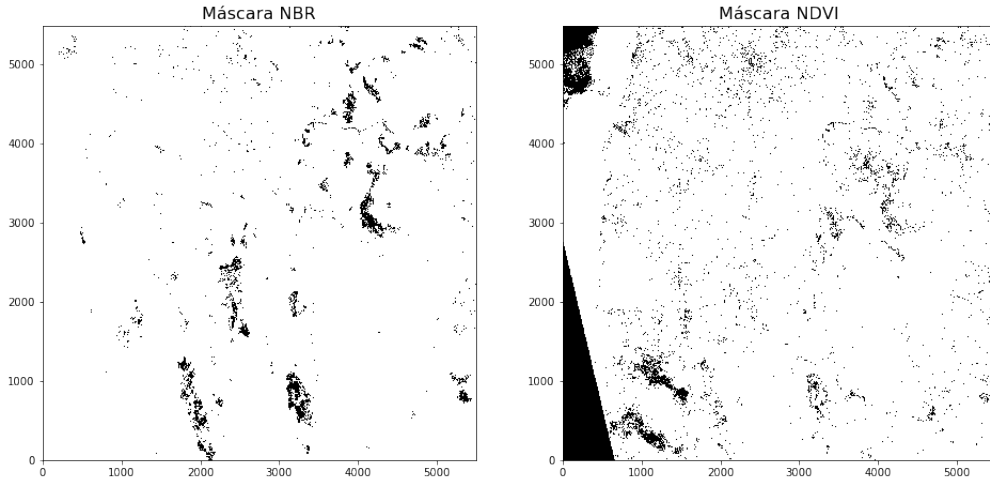


Figura 11: Máscaras originales que van a servir para comparar los resultados de los tres modelos usados para la detección del área incendiada de Galicia.

Los píxeles identificados como incendio en cada máscara:

$N_{nbr}$	$N_{ndvi}$
670344	2054043

Tabla 5: Número de píxeles incendio de la máscara original NBR y de la máscara original NDVI de Galicia.

Al igual que en el caso de Encinedo, la máscara incendio con el NBR es más limpia que la máscara con el NDVI. La diferencia entre número de píxeles de las máscaras de Galicia es notablemente mayor. Esto se debe a que, en el caso NDVI, se está identificando equívocamente zona de agua con incendio. Esta clasificación puede ser errónea debido al valor umbral del NDVI, se ha tomado el mismo valor pero podría ser distinto que el caso de Encinedo. Como ya se ha comentado, los valores del NDVI han de adaptarse a cada condición. Aunque esto parezca ser un problema, a la hora de comparar la máscara original NDVI con las obtenidas con los modelos, si estas últimas también identifican la zona agua como incendio, los resultados de los parámetros estadísticos son igualmente válidos que en el caso NBR.

#### 4.1.3. Resta de Imágenes: Encinedo

##### NBR

Se ha utilizado como imagen previa al incendio la propia al día 18/08/2017 y como posterior la del día 02/09/2017. Se recuerda que este incendio comienza el 21/08/2017.

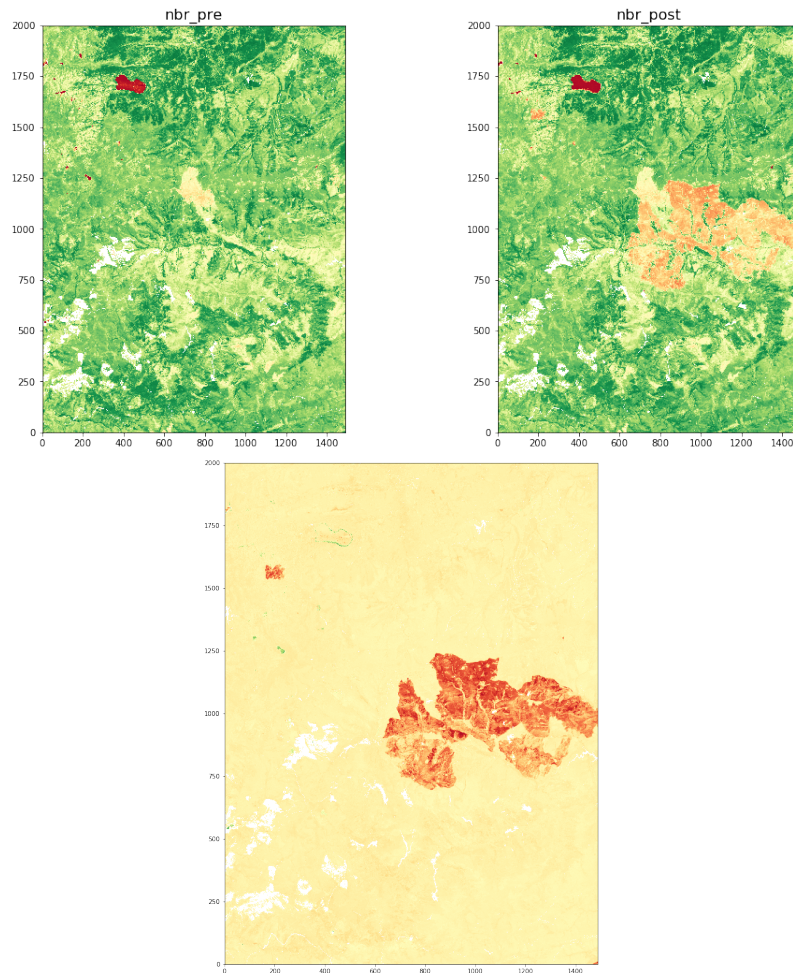


Figura 12: Imágenes utilizadas para el método de resta de imágenes sobre la zona de Encinedo (imagen superior). Se muestra la imagen pre incendio y la imagen post incendio, obtenidas para el índice NBR. Y el resultado de la resta de las imágenes (imagen inferior).

La matriz de confusión:

	Positivo	Negativo
Positivo	182208	51845
Negativo	10220	2735727

Figura 13: Matriz de confusión para el modelo de resta de imágenes en el incendio de Encinedo para el índice NBR.

r	MCC	HAI/ha	S	P
0.9468	0.8480	7288	0.9468	0.7785

Tabla 6: Ratio  $r$ , índice de Mathews MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método de resta de imágenes, utilizando datos del NBR en Encinedo.

La máscara y los parámetros que se han obtenido son satisfactorios.

## NDVI

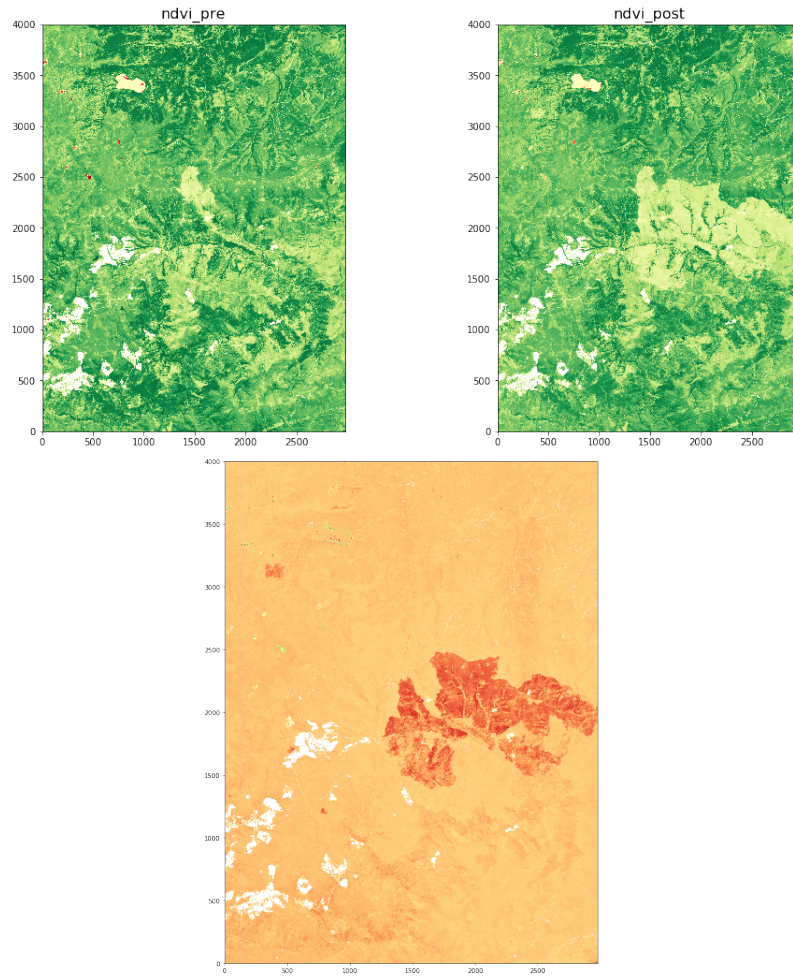


Figura 14: Imágenes utilizadas para el método de resta de imágenes sobre la zona de Encinedo, con los datos del índice NDVI (imagen superior). Se muestra la imagen pre incendio y la imagen post incendio. Y el resultado de la resta de las imágenes (imagen inferior).

La matriz de confusión para este caso:

	Positivo	Negativo
Positivo	108662	87950
Negativo	88236	2695152

Figura 15: Matriz de confusión para el modelo de resta de imágenes con datos del índice NDVI sobre el incendio de Encinedo.

r	MCC	HAI/ha	S	P
0.5518	0.5206	4346	0.5527	0.7785

Tabla 7: Ratio  $r$ , índice de Mathews MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método de resta de imágenes, utilizando datos del NDVI en Encinedo.

La máscara obtenida parece funcionar correctamente para identificar el perímetro de la superficie incendio. Pero existe un número de falsos positivos, lo suficientemente alto, como para provocar que los parámetros de la Tabla 7 no sean satisfactorios.

#### 4.1.4. Resta de imágenes: Galicia

##### NBR

Se ha utilizado como imagen previa al incendio la propia al día 12/10/2017 y como posterior la del día 22/10/2017. Se recuerda que este incendio comienza el 18/10/2017.



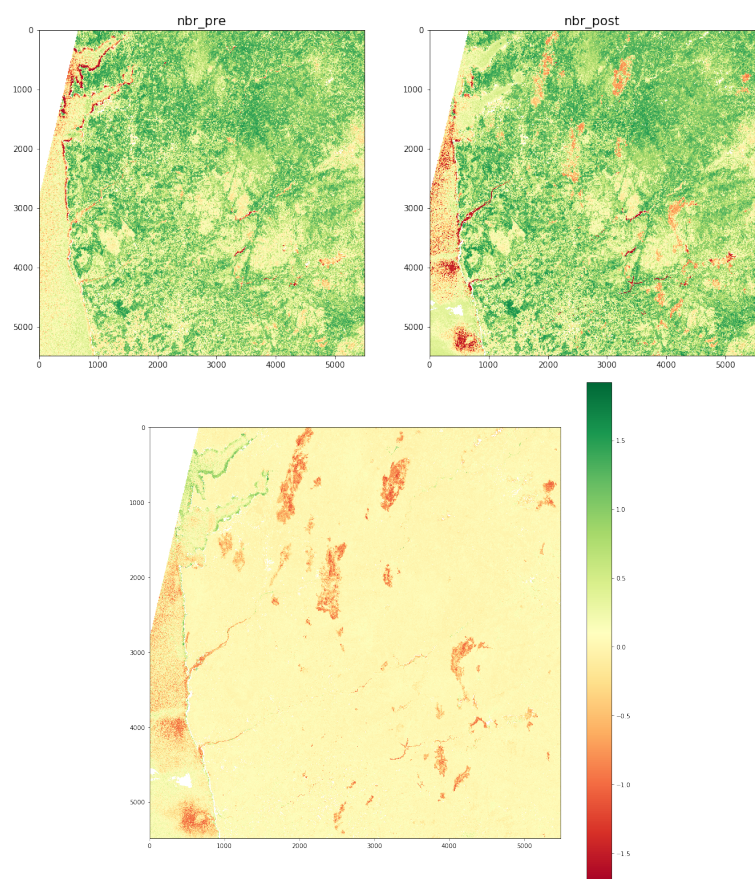


Figura 16: Imágenes utilizadas para el método de resta de imágenes sobre la zona de los incendios de Galicia, utilizando el índice NBR (imagen superior). Se muestra la imagen pre incendio y la imagen post incendio. Y resultado de la resta de las imágenes (imagen inferior).

La matriz de confusión del resultado de resta de imágenes en este caso es:

		Positivo	Negativo
0	Positivo	571377	98967
1	Negativo	12987	29456769

Figura 17: Matriz de confusión para el índice NBR de la resta de imágenes en Galicia.

r	MCC	HAI/ha	S	P
0.7814	0.5025	20954	0.9778	0.8527

Tabla 8: Ratio  $r$ , índice de Matthews MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método de resta de imágenes, utilizando datos del NBR en Galicia.

La máscara parece obtener correctamente los distintos incendios de Galicia. Los parámetros hallados son satisfactorios exceptuando el MCC, que queda por debajo de la sensibilidad  $S$  y la precisión  $P$ . En la máscara se puede ver que el problema de la identificación de la zona de agua con zona incendio se mantiene.

### NDVI

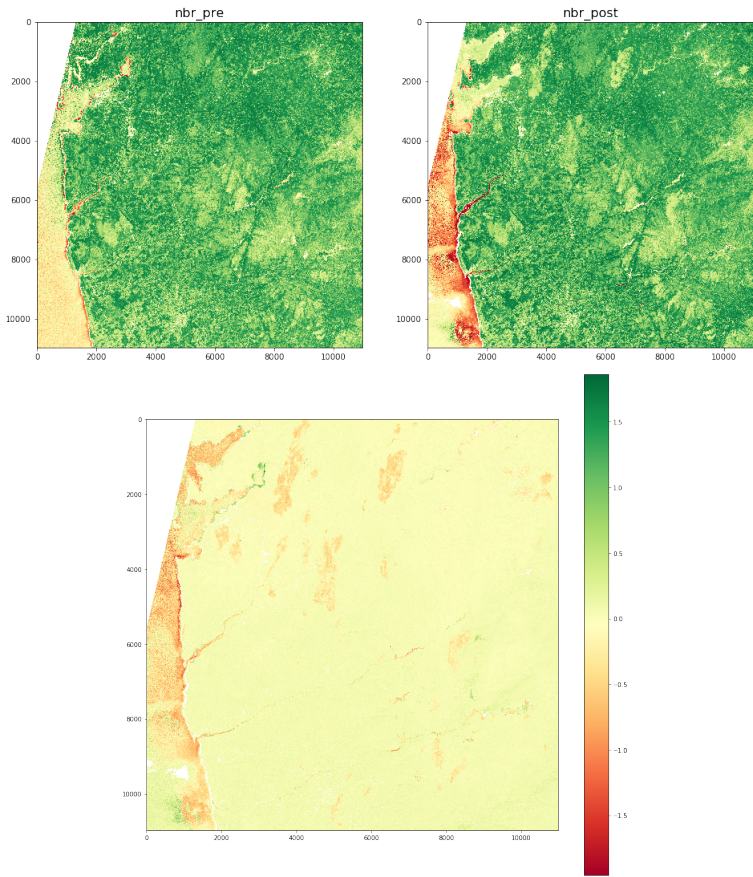


Figura 18: Imágenes utilizadas para el método de resta de imágenes sobre la zona de Galicia, con los datos del índice NDVI (imagen superior). Se muestra la imagen pre incendio y la imagen post incendio. Y el resultado de la resta de imágenes (imagen inferior).

La matriz de confusión:

	Positivo	Negativo
Positivo	445	1461976
Negativo	38	28677641

Figura 19: Matriz de confusión para el índice NDVI de la resta de imágenes en Galicia.

r	MCC	HAI/ha	S	P
0.0002	0.0162	20	0.9213	0.0003

Tabla 9: Ratio  $r$ , índice de Matthews MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método de resta de imágenes, utilizando datos del NDVI en Encinedo.

La máscara obtiene parte de los incendios de Galicia, pero la zona sureste queda levemente resaltada en comparación con el NBR. Los parámetros dictaminan que la efectividad, en este caso, no es buena. Se debe, principalmente, a la alta cantidad de falsos positivos.

#### 4.1.5. Método no supervisado, K-Means: Encinedo

El modelo K-Means utiliza en un mismo vector los datos del índice NDVI y NBR, por lo que no hay que separar los resultados según el índice. Antes de obtener la máscara con este método, se verifica el número de clústers óptimo. En la Tabla 4.1.5 se presentan los resultados del tiempo  $t$  que tarda el modelo en efectuarse, el ratio  $r$  y el índice MCC en función del número de clústers  $k$ .

$k$	$t/s$	$r$	MCC
2	30	0.9930	0.3840
3	68	0.9998	0.5019
4	70	0.9996	0.7776
5	105	0.9995	0.8906
6	141	0.9994	0.9041
7	175	0.9993	0.9288

Tabla 10: Tiempo  $t$  en segundos que tarda en efectuarse el método no supervisado, ratio  $r$  de píxeles frente a la máscara original e índice MCC para cada número  $k$  de clusters.

En la Tabla 10 se ha querido añadir el índice MCC debido a que el ratio  $r$  apenas varía con el número de clústers. En cambio, este índice utilizado para determinar la eficacia de los métodos, sí que permite determinar qué número  $k$  es el más eficiente.

Se puede considerar como mejor resultado en función del tiempo, el ratio y el MCC, el dado para  $k = 5$  clusters. Para visualizar mejor el resultado del K-Means, como lo que nos interesa es el cluster identificado como incendio, se presentan en la Figura 20 la imagen del K-Means con todos los clusters y la imagen que compara la máscara original y la máscara predicha.

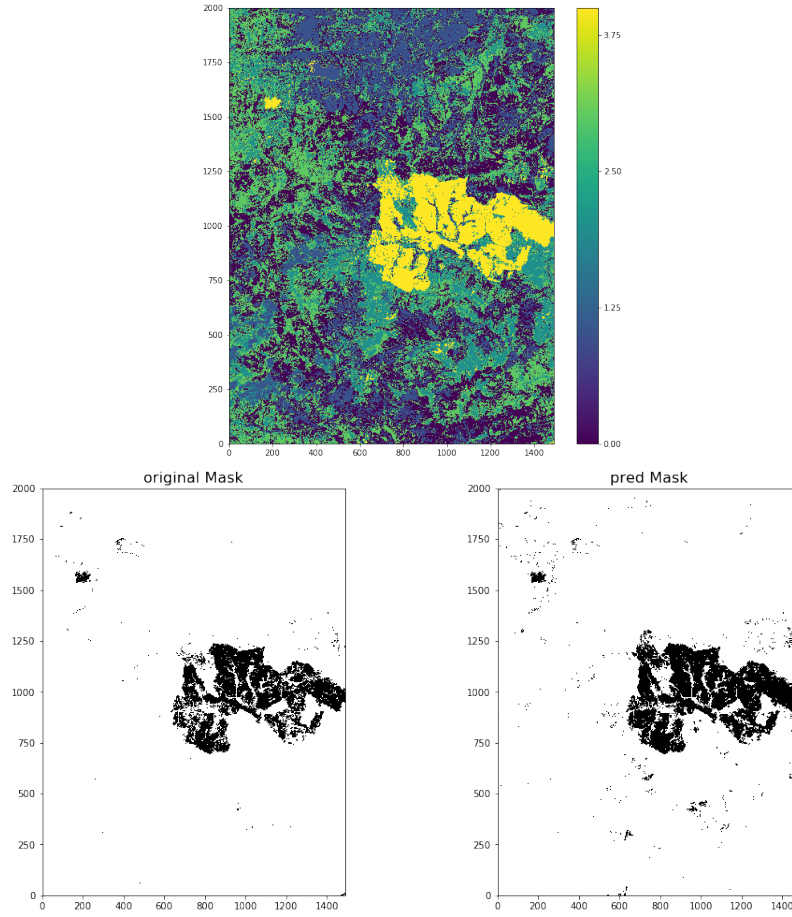


Figura 20: Resultado del modelo K-Means con 5 clústers (imagen superior) y comparación entre la máscara original (inferior izquierda) y la máscara del clúster incendio (inferior derecha), para el incendio de Encinedo.

Su matriz de confusión es:

	Positivo	Negativo
Positivo	192337	47815
Negativo	91	2739757

Figura 21: Matriz de confusión para el método K-Means en Encinedo.

r	MCC	HAI/ha	S	P
0.9995	0.8906	7694	0.9995	0.8001

Tabla 11: Ratio  $r$ , índice MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método no supervisado en Encinedo.

La comparación entre la máscara original y la máscara obtenida, demuestra que este método es notablemente preciso. Si bien es cierto que la máscara incendio

presenta píxeles sueltos, lo que implica que los parámetros MCC y  $P$  no sean tan buenos como el ratio  $r$  y la sensibilidad  $S$ .

#### 4.1.6. Método no supervisado, K-Means: Galicia

$k$	$t/s$	$r$	MCC
2	350	1.0	0.1669
3	863	0.9982	0.2489
4	1030	0.9946	0.2842
5	1455	0.9806	0.3617
6	1605	0.9955	0.4873

Tabla 12: Tiempo en segundos que tarda en efectuarse el método no supervisado y el ratio de píxeles frente a la máscara original para cada número  $k$  de clusters en Galicia.

Se puede considerar como mejor resultado en función del tiempo, el ratio y el MCC, el dado para  $k = 6$  clusters. Se presentan en la Figura 22 la imagen del K-Means con todos los clusters y la imagen que compara la máscara original y la máscara predicha.

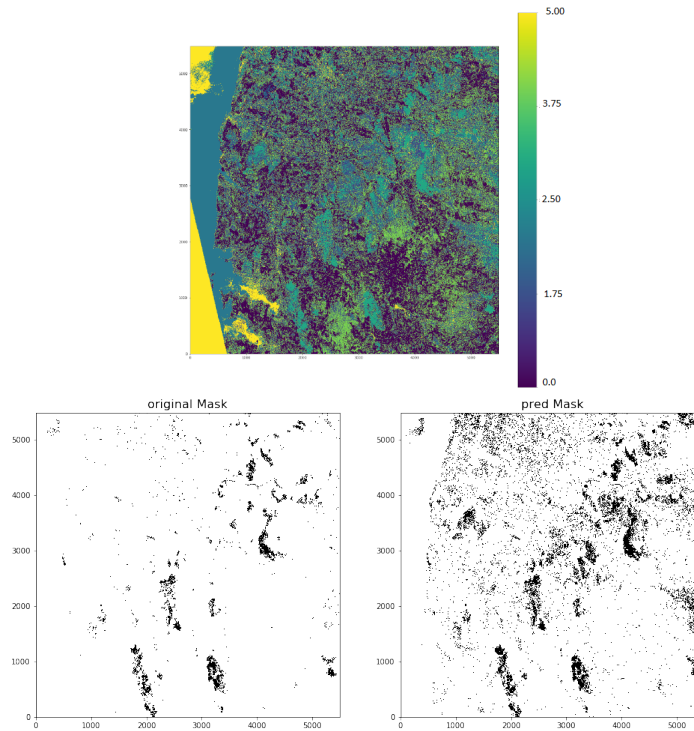


Figura 22: Resultado de KMeans para 6 clusters en Galicia (imagen superior) y comparación entre la máscara original (inferior izquierda) con la máscara obtenida a través de este método (inferior derecha).

Su matriz de confusión es:

	Positivo	Negativo
Positivo	667351	1837615
Negativo	2993	27632141

Figura 23: Matriz de confusión para el método K-Means en Galicia.

r	MCC	HAI/ha	S	P
0.3248	0.4873	26694	0.9955	0.2664

Tabla 13: Ratio  $r$ , índice MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método no supervisado en Galicia.

La máscara incendio, aunque clasifica correctamente los incendios, muestra mucho ruido. El número de falsos positivos es importante, como indican los parámetros MCC,  $P$  y  $r$ , que permiten verificar que este método no funciona del todo bien en Galicia.

#### 4.1.7. Método supervisado, Logistic Regression: Encinedo

El método de Regresión Logística se ha utilizado empleando datos del índice NBR y del índice NDVI por separado. Se presentan los resultados para ambos índices en la Tabla 4.1.7 donde aparece el tiempo que el modelo tarda en entrenar los resultados  $t$  así como el ratio  $r$  entre los píxeles de la máscara original.

Proporción entrenada	$t_{nbr}/s$	$r_{nbr}$	$t_{ndvi}/s$	$r_{ndvi}$
0.2	12	0.9885	10	0.4702
0.3	10	0.9886	9	0.4662
0.4	8	0.9887	8	0.4622
0.5	7	0.9888	7	0.4560
0.6	6	0.9887	5	0.4438
0.7	4	0.9889	4	0.4303
0.8	2	0.9885	2	0.3978

Tabla 14: Incendio de Encinedo. Tiempo en segundos que tarda en entrenarse el método supervisado y el ratio de píxeles frente a la máscara original para cada proporción de datos con los que se entrena el modelo, para el índice NBR y el índice NDVI.

### Índice NBR

Según los valores de la Tabla 14 los mejores resultados se obtienen entrenando un 70 %. Se presenta en la Figura 24 la imagen predicha de la máscara del incendio así como la imagen de la máscara original para esta proporción de datos a entrenar.

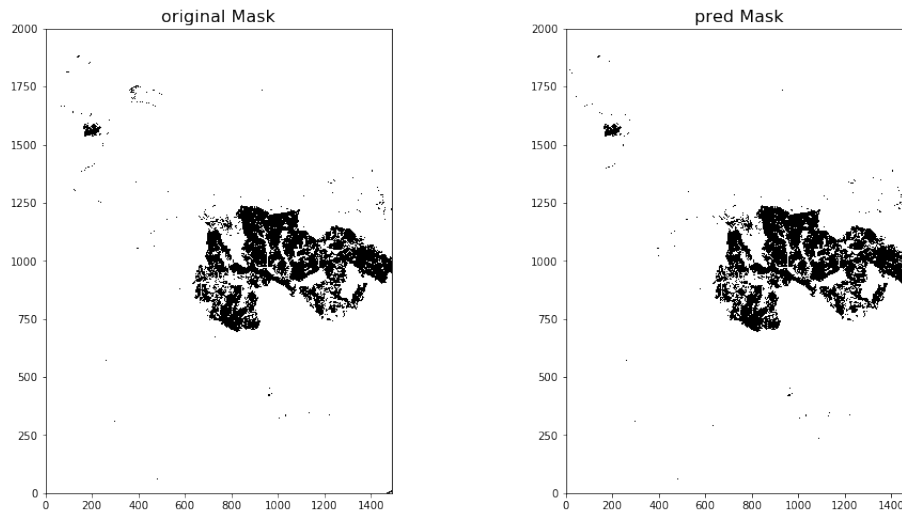


Figura 24: Resultado gráfico del método supervisado con el índice NBR en el área de Encinedo tras el incendio con una proporción de datos a entrenar de 0,7. Se muestra la máscara con la que se ha entrenado el modelo (original mask) y el resultado de este (pred mask).

La matriz de confusión:

	Positivo	Negativo
Positivo	190288	2140
Negativo	3022	2784550

Figura 25: Matriz de confusión para el índice NBR en el método supervisado.

r	MCC	HAI/ha	S	P
0.9889	0.9857	7611	0.9843	0.9890

Tabla 15: Ratio  $r$ , índice MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método supervisado con datos NBR en Encinedo.

La máscara incendio hallada, si analizamos la Figura 24, es prácticamente igual que la máscara original. Los parámetros de la Tabla 15 permiten dictaminar que este caso es sumamente eficaz.

## Índice NDVI

Volviendo a observar los datos de la Tabla 14 se verifica que, para el NDVI, los mejores resultados se obtienen entrenando un 40 %.

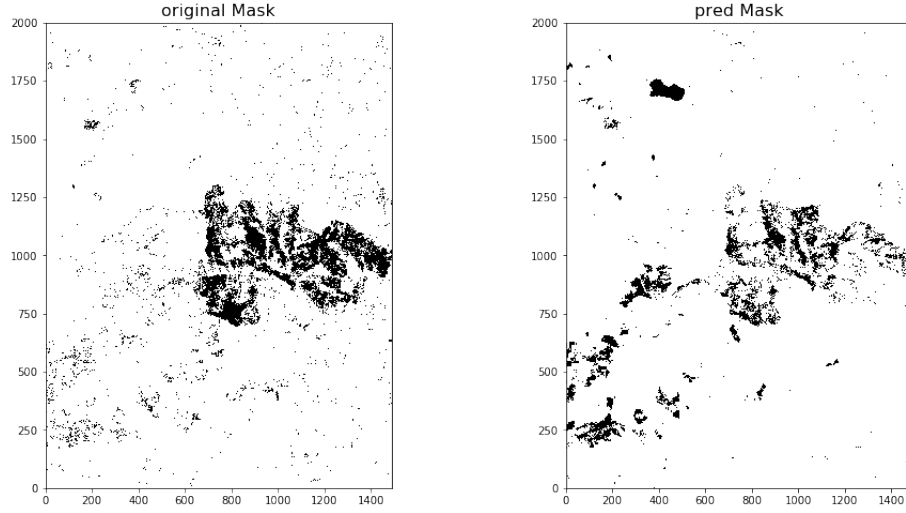


Figura 26: Resultado gráfico del método supervisado con el índice NDVI en el área de Encinedo tras el incendio con una proporción de datos a entrenar de 0,4. Se muestra la máscara con la que se ha entrenado el modelo (original mask) y el resultado de este (pred mask).

	Positivo	Negativo
Positivo	91005	105893
Negativo	66923	2716179

Figura 27: Matriz de confusión para el método supervisado en el incendio de Encinedo para el índice NDVI.

r	MCC	HAI/ha	S	P
0.4622	0.4858	910	0.5762	0.0324

Tabla 16: Ratio  $r$ , índice MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método supervisado, utilizando el NDVI, en Encinedo.

La máscara incendio aquí obtenida presenta bastante ruido. Identificando áreas incendio donde no corresponde. Los parámetros de la Tabla 16 así lo demuestran y en concreto, la precisión  $P$ , clarifica esta idea.



#### 4.1.8. Método supervisado, Logistic Regression: Galicia

Proporción entrenada	$t_{nbr}/s$	$r_{nbr}$	$t_{ndvi}/s$	$r_{ndvi}$
0.2	705	0.8705	606	0.0166
0.3	512	0.8675	569	0.0166
0.4	445	0.8635	481	0.0165
0.5	368	0.8587	422	0.0167
0.6	301	0.8523	344	0.0167
0.7	221	0.8431	233	0.0166
0.8	164	0.8286	137	0.0166

Tabla 17: Tiempo en segundos que tarda en efectuarse el método supervisado y el ratio de píxeles frente a la máscara original para cada proporción de datos con los que se entrena el modelo, para el índice NBR y el índice NDVI en el incendio de Galicia.

#### NBR

Dentro de los resultados de la Tabla 17 se ha considerado que la proporción óptima de datos a entrenar para el caso del índice NBR es 0,6.

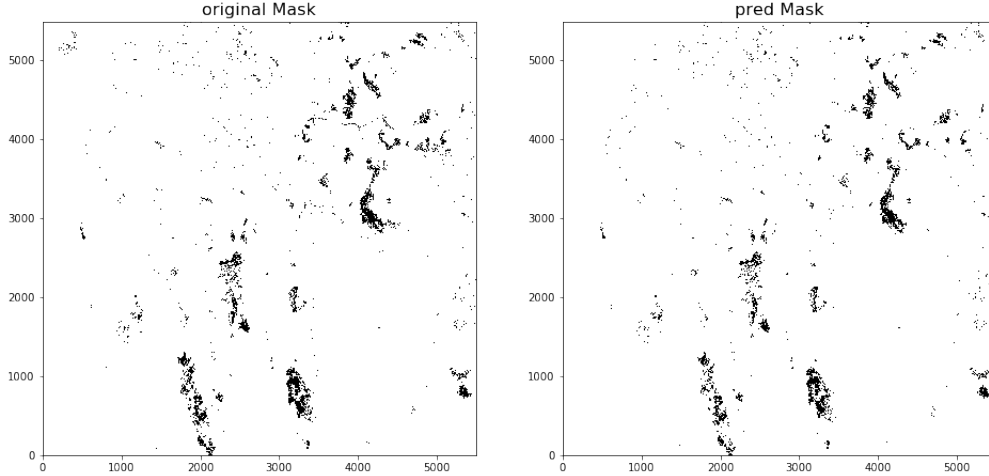


Figura 28: Resultado gráfico del método supervisado con el índice NBR en el área de Galicia tras los incendios con una proporción de datos a entrenar de 0,6. Se muestra la máscara con la que se ha entrenado el modelo (original mask) y el resultado de este (pred mask).

La matriz de confusión para este caso es:

		Positivo	Negativo
0	Positivo	571377	98967
1	Negativo	12987	29456769

Figura 29: Matriz de confusión para el índice NBR en el método supervisado en Galicia.

r	MCC	HAI/ha	S	P
0.8523	0.9111	22855	0.9778	0.8524

Tabla 18: Ratio  $r$ , índice MCC, hectáreas incendiadas HAI, sensibilidad S y precisión P, según el método supervisado, utilizando el NBR, en Galicia.

La máscara hallada muestra alta precisión comparándola con la original, siendo prácticamente iguales. Aunque el número de falsos positivos hacen que los parámetros de este caso sean algo peores que los del caso de Encinedo.

## NDVI

Los resultados usando el índice NDVI en Galicia no son efectivos. El ratio independientemente de la proporción de datos utilizados es cercano a 0 y el MCC es negativo. No se tienen en cuenta los resultados, se muestra el resultado gráfico en la Figura 29 y su matriz de confusión en la Figura 31 para verificar su ineficacia.

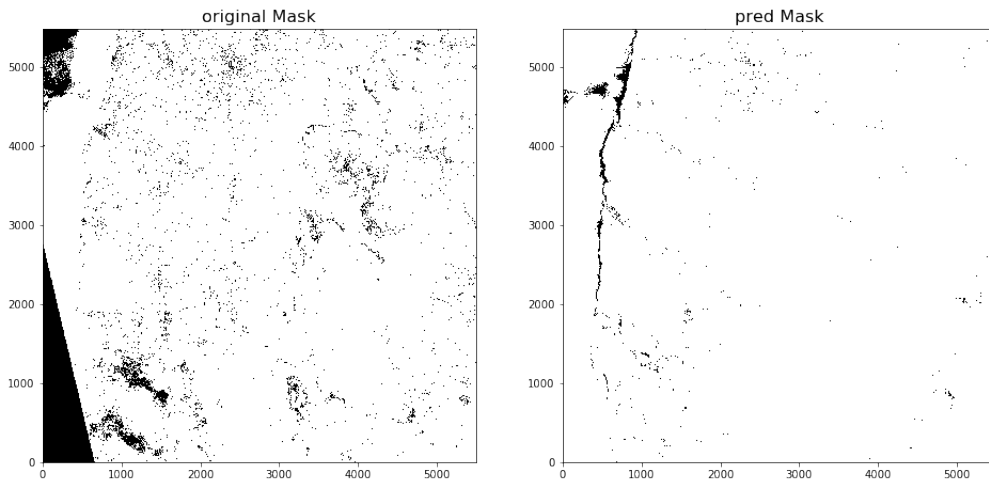


Figura 30: Resultado gráfico del método supervisado con el índice NDVI en el área de Galicia tras los incendios con una proporción de datos a entrenar de 0,7. Se muestra la máscara con la que se ha entrenado el modelo (original mask) y el resultado de este (pred mask).

		Positivo	Negativo
0	Positivo	11190	2042853
1	Negativo	241968	27844089

Figura 31: Matriz de confusión para el índice NDVI en el método supervisado en Galicia.

r	MCC	HAI/ha	S	P
0.005	448	344	0.044	0.0054

Tabla 19: Ratio  $r$ , índice MCC, hectáreas incendiadas HAI, sensibilidad  $S$  y precisión  $P$ , según el método supervisado, utilizando el NDVI, en Galicia.

La máscara y los parámetros que se han hallado son defectuosos.

## 4.2. Análisis de resultados de detección de incendios

En la Tabla 4.2 se resume los resultados principales del incendio de Encinedo:

Modelo	índices	r	HAI/ha	MCC	t/s	S	P
Resta de Imágenes	NBR	0.9468	7288	0.8480	2	0.9468	0.7785
	NDVI	0.5318	4346	0.5209	2	0.5527	0.5518
No supervisado		0.9995	7693	0.8906	105	0.9995	0.8001
Supervisado	NBR	0.9889	7611	0.9857	4	0.9843	0.9890
	NDVI	0.4622	910	0.4858	8	0.5762	0.0324

Tabla 20: Resultados principales de los tres modelos utilizados para la detección de incendios en Encinedo.

Como mejor modelo se tiene el método supervisado, la regresión logística, con datos del índice NBR.

Si se compara el ratio obtenido con el método no supervisado y el obtenido con el método supervisado, se tiene que el primero es un 1 % mejor, pero el MCC es un 11 % menor, al igual que la precisión  $P$  que lo es un 19 %. Esto implica que la máscara final contiene un mayor porcentaje de píxeles erróneamente clasificados como incendio. Además, el tiempo que tarda el método supervisado es tan solo de 4 s, mientras que el del método no supervisado es de 105 s. El método de resta de imágenes es el que peor resultado tiene de los tres en el ratio, en el MCC, en la sensibilidad y en la precisión. Si bien es cierto que es el más rápido de todos, aunque la diferencia con el método supervisado es pequeña. Otro resultado importante es

la diferencia entre los índices NDVI y NBR. Esta diferencia puede deberse a que el índice NBR es creado con aplicación directa a superficies quemadas, mientras que el índice NDVI, como se ha comentado, es útil en aspectos más generales sobre la vegetación. Otra razón es la elección del valor umbral de cada índice. En el caso del índice NBR está bien determinado dicho valor, a nivel bibliográfico, y es el más utilizado en la mayoría de casos. En cambio, en el índice NDVI varían más sus intervalos establecidos dependiendo del caso en el que se está aplicando.

En la Tabla 21 se presentan los resultados principales de los incendios de Galicia:

Modelo	índices	r	HAI/ha	MCC	t/s	S	P
Resta de Imágenes	NBR	0.7814	20954	0.5025	4	0.9778	0.8527
	NDVI	0.0002	20	0.0016	6	0.9213	0.0003
No supervisado		0.9955	26694	0.4873	1605	0.9955	0.2664
Supervisado	NBR	0.8583	22853	0.9111	301	0.9778	0.8524
	NDVI	0.005	344	-0.0087	350	0.044	0.0054

Tabla 21: Resultados principales de los tres modelos utilizados para la detección de incendios en Galicia.

Los resultados cumplen el mismo comportamiento que en el caso de Encinedo. El método supervisado para el índice NBR es el mejor de los tres utilizados. El MCC, el tiempo que tarda en efectuarse la clasificación y la precisión, son notablemente mejores que el caso no supervisado. Aunque el tiempo sea superior frente a la resta de imágenes, los valores de los parámetros utilizados marcan la diferencia a favor del método supervisado. También se repite el hecho de que el NBR funciona mejor que el NDVI, en este caso de Galicia la diferencia es mayor que en Encinedo.

Si comparamos los resultados del incendio de Encinedo y de los incendios de Galicia, se determina que para el primero la efectividad de los métodos e índices en general, y de la regresión logística con el NBR en particular son mejores. Esto puede deberse a varios factores, pero principalmente se puede destacar la cantidad de datos de un caso y del otro. El incendio de Encinedo está localizado en un área más pequeña que los de Galicia, por lo que los modelos pueden procesar más rápido y con menos errores los datos pertinentes. Además, en el caso de Galicia no se está analizando un incendio en concreto, aparecen varias zonas afectadas, lo que complica la clasificación ya que los píxeles incendio están más dispersos.

En conclusión, el mejor método es el supervisado usando datos del NBR. Tanto para el caso de Encinedo como de Galicia funciona satisfactoriamente. Este modelo tiene una ventaja importante y es que, una vez entrenado no hace falta volver a hacerlo. Es decir, se puede ejecutar el modelo para nuevos datos y este los clasifica

automáticamente, lo que es bastante útil para la siguiente sección de resultados donde analizaremos la evolución temporal de los incendios.

### 4.3. Evolución temporal de la superficie del incendio

#### 4.3.1. Encinedo

En la Figura 32 se presentan las imágenes del NBR y en la Figura 34 las del NDVI para cuatro distintas fechas, donde se aprecia el cambio en la zona.

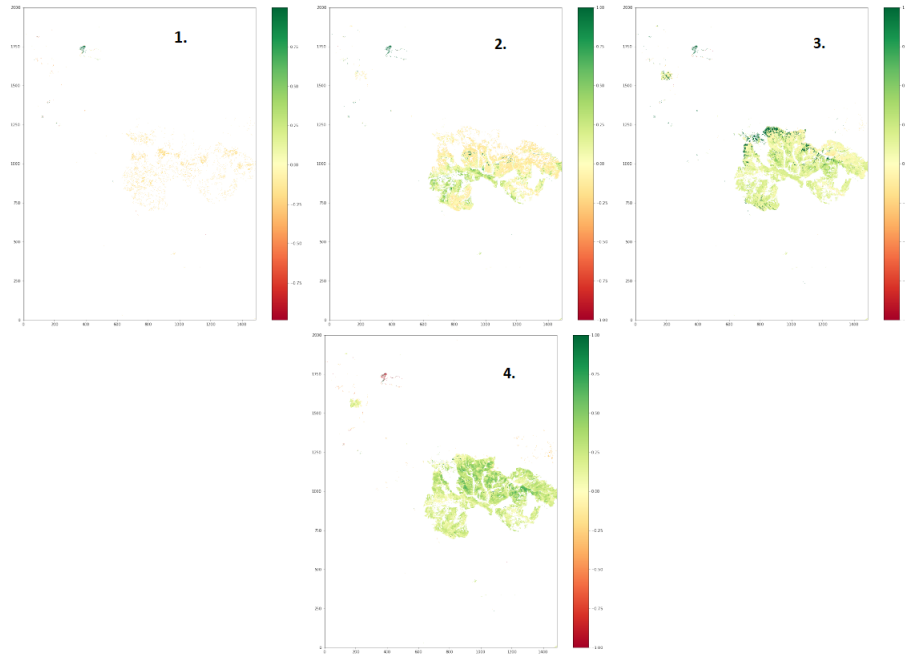


Figura 32: Secuencia de imágenes del índice NBR en el área afectada por el incendio de Encinedo. La primera imagen (1) es del 2017/10, la segunda (2) del 2018/06, la tercera (3) del 2019/02 y la cuarta (4) del 2020/02.

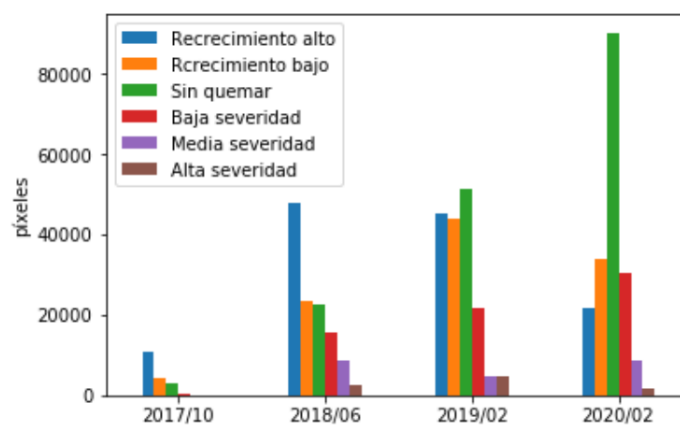


Figura 33: Número de píxeles, de las imágenes de la Figura 32, para cada intervalo de valores pertinentes al índice dNBR de la Tabla 3.

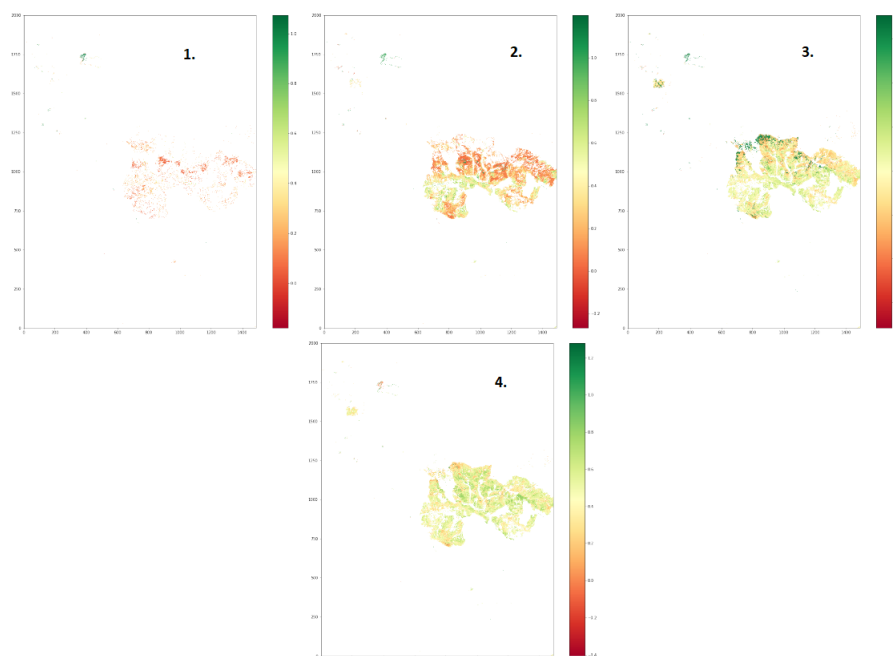


Figura 34: Secuencia de imágenes del índice NDVI en el área afectada por el incendio de Encinedo. La primera imagen (1) es del 2017/10, la segunda (2) del 2018/06, la tercera (3) del 2019/02 y la cuarta (4) del 2020/02.

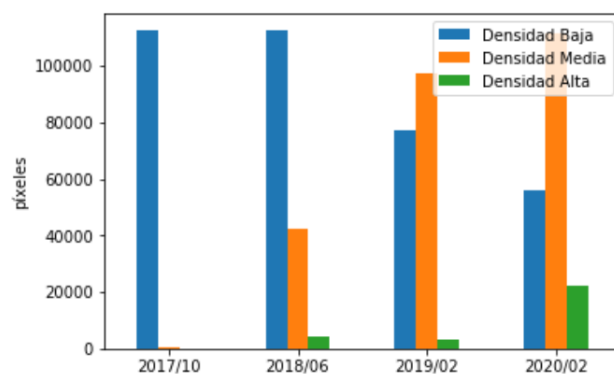


Figura 35: Número de píxeles, de las imágenes de la Figura 34, para cada intervalo de valores pertinentes al índice NDVI de la Tabla 2.

#### 4.3.2. Galicia

En la Figura 36 se presentan las imágenes del NBR y en la Figura 38 del NDVI para cuatro distintas fechas, donde se aprecia el cambio en la zona.

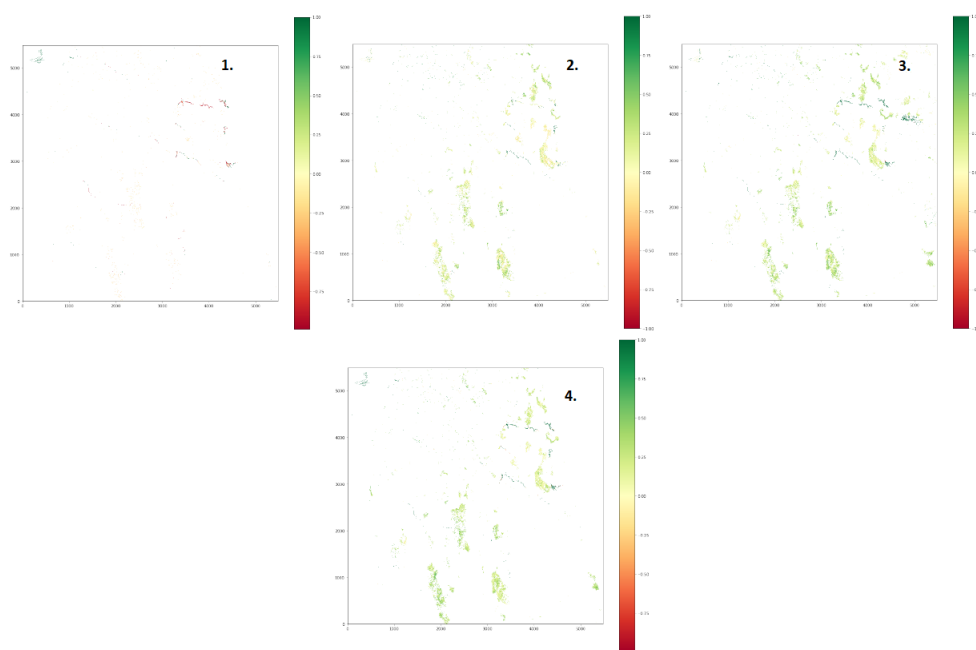


Figura 36: Secuencia de imágenes del índice NBR en el área afectada por el incendio de Galicia. La primera imagen (1) es del 2017/11, la segunda (2) del 2018/08, la tercera (3) del 2019/08 y la cuarta (4) del 2020/02.

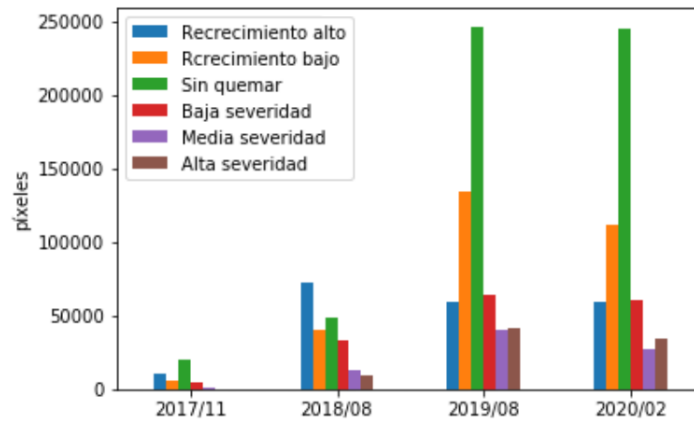


Figura 37: Número de píxeles, de las imágenes de la Figura 36, para cada intervalo de valores pertinentes al índice dNBR de la Tabla 3.

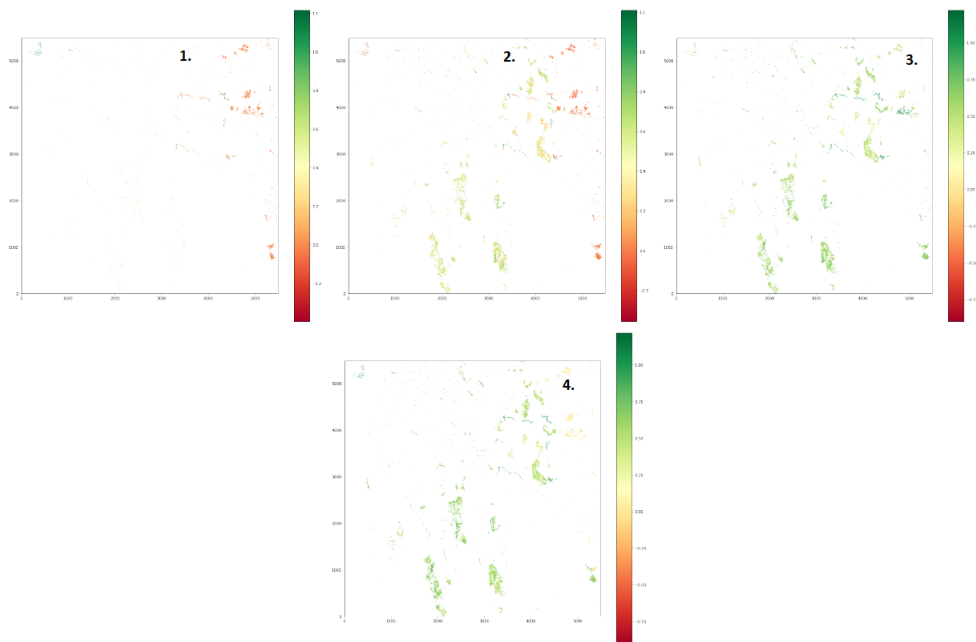


Figura 38: Secuencia de imágenes del índice NDVI en el área afectada por el incendio de Galicia. La primera imagen (1) es del 2017/11, la segunda (2) del 2018/08, la tercera (3) del 2019/08 y la cuarta (4) del 2020/02.



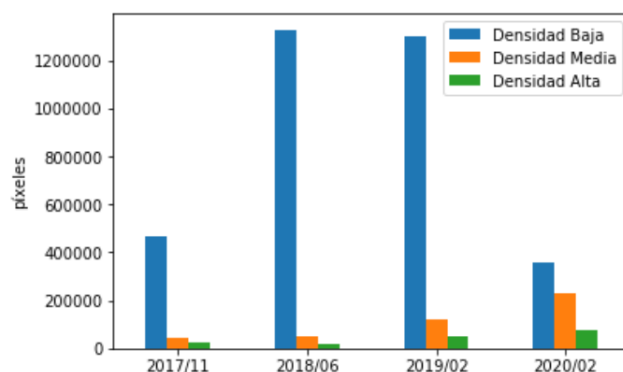


Figura 39: Número de píxeles, de las imágenes de la Figura 38 para cada intervalo de valores pertinentes al índice NDVI de la Tabla 2.

#### 4.4. Análisis de resultados del avance temporal del incendio

Para los valores del dNBR, el intervalo que más crece tanto en Encinedo como en Galicia es zona sin quemar, la frontera entre zona de recuperación y zona de severidad de incendio. Entre la primera y segunda imagen, el mayor crecimiento se da para los intervalos negativos, que son zonas en proceso de recuperación. Para la tercera y cuarta imagen, el intervalo de la zona de sin quemar mantiene su crecimiento, mientras el resto crece o disminuye levemente, lo que se traduce en predominancia de este intervalo de valores.

Para los valores del NDVI el intervalo que más crece es el densidad de vegetación media. Al principio, los píxeles más predominantes son los del intervalo de densidad baja pero a medida que avanza el tiempo estos decaen mientras aumentan los píxeles de los otros intervalos. Los píxeles del intervalo propio a las zonas de vegetación muy densa, son prácticamente inexistentes al principio y comienza su crecimiento en la última imagen. Proyectando esta tendencia en fechas futuras, se verificaría la recuperación del vegetación afectada tras el incendio.

Según los valores del dNBR, los incendios de Galicia se recuperan más rápido que el incendio de León. El motivo de esta idea es que los píxeles de los intervalos de zonas de recuperación, tienen un ratio de crecimiento mayor en Galicia que en Encinedo. Una forma de explicar esta diferencia está en la vegetación. Utilizando el portal de datos de biodiversidad GBIF, se puede buscar las especies vegetales de un área. En la Figura 40 se presentan las imágenes y el registro estadístico de la vegetación de Encinedo y de uno de los incendios de Galicia.

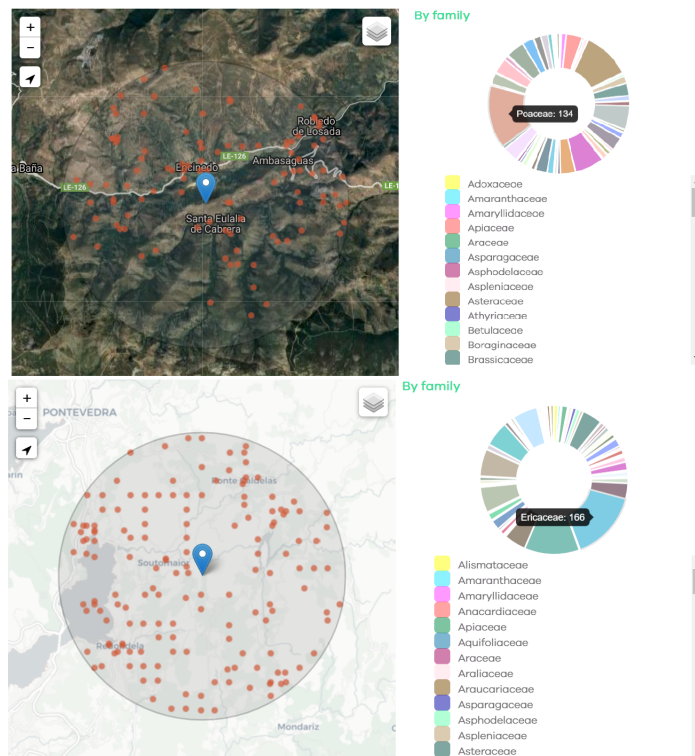


Figura 40: Área del incendio de Encinedo y estadísticas de las familias de vegetación en ella (imagen izquierda). Lo mismo pero para uno de los incendios de Galicia (imagen derecha).

Siguiendo las estadísticas, las tres familias de vegetación predominante de Encinedo son: Poaceae (111), Asteraceae (96) y Fabaceae (52). Mientras que las de Galicia son: Ericaceae (150), Fabaceae (135) y Myrtaceae (71). Según [40] las especies de la familia Myrtaceae se adaptan satisfactoriamente a los incendios. Además, la familia Fabaceae que es predominante en ambas zonas, es invasora y se reproduce fácilmente tras un incendio, favoreciendo su crecimiento. Según la estadística del portal GBIF, Galicia tiene el doble de especies Fabaceae que Encinedo, lo que sumado a la característica de la familia Myrtaceae sirve para determinar que la recuperación en Galicia, según la vegetación, es más rápida que en Encinedo.

## 5. Conclusiones

Se han presentado los resultados de tres técnicas de análisis de datos para detectar y clasificar los incendios de 2017 de Encinedo y Galicia.

Se ha determinado que, de los índices NBR y NDVI, el mejor para nuestro propósito es el NBR. Superando al NDVI, tanto en Encinedo como en Galicia, en los parámetros  $r$ ,  $MCC$ ,  $S$  y  $P$ .

De los tres modelos, se ha comprobado que el método supervisado con el índice NBR, modelo de regresión logística, es el más eficiente. Los resultados en Encinedo utilizando el índice NBR y el método supervisado permiten concluir que esta técnica es notablemente eficaz para este caso, obteniéndose que  $r = 0,9889$ ,  $MCC = 0,9857$ ,  $S = 0,9843$  y  $P = 0,9890$ . En el caso de Galicia, también es el método supervisado con el índice NBR el más eficaz, aunque menos que para Encinedo, con resultados de  $r = 0,8583$ ,  $MCC = 0,9111$ ,  $S = 0,9778$  y  $P = 0,8524$ .

Se concluye que los resultados de Encinedo son mejores que los resultados de Galicia. Se ha estimado que el motivo es la diferencia de proporción de datos que hay que procesar para cada caso. Siendo menos y estando más localizados en Encinedo, lo que facilita su tratamiento. Por lo tanto se puede predecir en futuros casos que nuestra técnica funciona mejor cuanto más localizado esté el incendio. Si el incendio o los incendios están dispersos en un área grande se pierde eficacia.

Se ha utilizado el método supervisado con datos del índice NBR para obtener la evolución temporal de la superficie afectada por los incendios. Para su análisis se han utilizado los índices NDVI, NBR y dNBR. Se ha determinado que la zona que mejor recuperación tiene es la de Galicia y se ha dado una explicación a esta diferencia en la regeneración desde el punto de vista de la vegetación de cada zona.

En conclusión, se ha comprobado que mediante el modelo de Regresión Logística, propio al método de aprendizaje supervisado, y el índice de calcinación NBR, se puede detectar y monitorizar un incendio forestal.

## Referencias

- [1] Leblon, B. (2005). *Monitoring Forest Fire Danger with Remote Sensing*. Natural Hazards, 35(3), 343–359.
- [2] Glosario: Incendio forestal (2018). Recuperado 11 de mayo de 2020, de AEMET website: [https://meteoglosario.aemet.es/es/termino/342\\_incendio-forestal](https://meteoglosario.aemet.es/es/termino/342_incendio-forestal).
- [3] Adrián Cardil Forradellas (2013). *Víctimas en incendios forestales en España en el periodo 1980-2010*. 6o congreso forestal español. Escuela técnica superior de ingeniería agraria, Universidad de Lleida.
- [4] Rittmaster, Robyn Adamowicz, Wiktor Amiro, B. Pelletier, Rick. (2006). *Economic analysis of health effects from forest fires*. Canadian Journal of Forest Research-revue Canadienne De Recherche Forestiere - CAN J FOREST RES. 36. 868-877.
- [5] Zhang, Yanyan Obrist, Daniel Zielinska, Barbara Gertler, Alan. (2013). *Particulate emissions from different types of biomass burning*. Atmospheric Environment. 72. 27-35.
- [6] Bond, William Keeley, Jon. (2005). *Fire as a Global ‘Herbivore’: The Ecology and Evolution of Flammable Ecosystems*. Trends in ecology evolution. 20. 387-94.
- [7] Nasi, Robert Dennis, Rona Meijaard, Erik Applegate, Grahame Moore, Peter. (2002). *Forest fire and biological diversity*. Unasylva 209. 53.
- [8] Zedler, Paul Gautier, Clayton McMaster, Gregory. (1983). *Vegetation Change in Response to Extreme Events: The Effect of a Short Interval between Fires in California Chaparral and Coastal Scrub*. Ecology. 64. 809.
- [9] Benavides Ballesteros, H.; León Aristazabal, G. *Información técnica sobre gases de efecto invernadero y el cambio climático*. Nota técnica del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). 2007.
- [10] van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S. *Global fire emissions estimates during 1997–2016*. Earth Syst. Sci. Data, 9, 697–720.
- [11] Cruz-Núñez, Xochitl Villers, Lourdes Gay Garcia, Carlos. (2013). *4.2 Incendios forestales, carbono negro y carbono orgánico en México: 2000 – 2012*. Estado actual del conocimiento del ciclo del carbono y sus interacciones en México: Síntesis al 2013. pp.596-601.
- [12] Interamerican Association For Environmental Defense, AIDA. (2013). *Contaminantes Climáticos de Vida Corta: una oportunidad para reducir emisiones*.

- [13] Informe clima e incendios forestales (2015). Recuperado 20 de abril de 2020 de WWF website: <https://www.wwf.es/?39717/Informe-Clima-e-Incendios-Forestales>.
- [14] Sousa, Pedro Trigo, Ricardo Pereira, Mário Bedia, Joaquín Gutiérrez, J.. (2015). *Different approaches to model future burnt area in the Iberian Peninsula*. Agricultural and Forest Meteorology. 202.
- [15] Ministerio de Agricultura y Pesca, Alimentación y Medioambiente. Gobierno de España. *Los incendios forestales en España*. 1 de enero - 31 de diciembre. Avance informativo. 2017.
- [16] Ministerio de Agricultura y Pesca, Alimentación y Medioambiente. Gobierno de España. *Los incendios forestales en España*. 1 de enero - 31 de diciembre. Avance informativo. 2017.
- [17] Navarrete, Edison Laubacher, Gerard. (2017). *APUNTES DE TELEDETECCIÓN: Teledetección Espacial y Procesamiento de Imágenes*.
- [18] Emilio Chuvieco. 2016. *Fundamentals of Satellite Remote Sensing: An Environmental Approach*, Second Edition (2nd. ed.). CRC Press, Inc., USA.
- [19] Aschbacher, J. (2017). *ESA's earth observation strategy and Copernicus*. In Satellite earth observations and their impact on society and policy (pp. 81-86). Springer, Singapore.
- [20] European Space Agency (ESA). *Sentinel-2, User Handbook*. Addison-Wesley, Reading, Massachusetts, 2015.
- [21] Jovanovska Kaplan, Gordana Avdan, Ugur. (2017). *Object-based water body extraction model using Sentinel-2 satellite imagery*. European Journal of Remote Sensing. 50. 137-143.
- [22] Hayt, W. H., Buck, J. A., Pedraza, C. R. C. (2006). *Teoría electromagnética*. McGraw-Hill.
- [23] Loudon, R. (2000). *The quantum theory of light*. OUP Oxford.
- [24] Agencia Espacial Europea, ESA. *Teledetección, a fondo* Recuperado el 18 de mayo de 2020, de ESA eduspace website: [http://www.esa.int/SPECIALS/Eduspace\\_ES/SEMZ3YD3GXF\\_0.html](http://www.esa.int/SPECIALS/Eduspace_ES/SEMZ3YD3GXF_0.html).
- [25] Francisco Sacristán Romero. 2006. *La Teledetección satelital y los sistemas de protección ambiental* Revista AquaTIC, No 24. Revista científica de la Sociedad Española de Acuicultura.
- [26] Emilio Chuvieco. 2016. *Fundamentals of Satellite Remote Sensing: An Environmental Approach*, Second Edition (2nd. ed.). CRC Press, Inc., USA.
- [27] Knipling, E. B. (1970). *Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation*. Remote sensing of environment, 1(3), 155-159.

- [28] Brent N, Holben (1986). *Characteristics of maximum-value composite images from temporal AVHRR data*. International Journal of Remote Sensing, 7:11, 1417-1434.
- [29] Parmar, Manan Shukla, Shital Kalubarme, Manik. (2019). *IMPACT OF CLIMATE CHANGE AND DROUGHT ANALYSIS ON AGRICULTURE IN SABARKANTHA DISTRICT USING GEOINFORMATICS TECHNOLOGY*. 6. 133 - 144.
- [30] Rizvi, Raza Yadav, R Singh, Ramesh Datt, Keshav Khan, I Dhyani, S. (2009). *Spectral Analysis of Remote Sensing Image for Assessment of Agroforestry Areas in Yamunanagar district of Haryana*.
- [31] Roy, D. P., Boschetti, L., Trigg, S. N. (2006). *Remote sensing of fire severity: assessing the performance of the normalized burn ratio*. IEEE Geoscience and Remote Sensing Letters, 3(1), 112-116.
- [32] Key, Carl.; Benson, Nate. *Landscape Assessment (LA) Sampling and Analysis Methods*. FIREMON: Fire Effects Monitoring and Inventory System 2005.
- [33] Athanasakis, George Psomiadis, Emmanouil Chatziantoniou, Andromachi. (2017). *High-resolution Earth observation data and spatial analysis for burn severity evaluation and post-fire effects assessment in the Island of Chios, Greece*. 61.
- [34] Jain, A. K. (2010). *Data clustering: 50 years beyond K-means*. Pattern recognition letters, 31(8), 651-666.
- [35] Stuart Russell and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach* (3rd. ed.). Prentice Hall Press, USA. pp 725-727.
- [36] Confusion matrix. (2020). Recuperado 10 de mayo de 2020, de Wikipedia website: [https://en.wikipedia.org/wiki/Confusion\\_matrix](https://en.wikipedia.org/wiki/Confusion_matrix).
- [37] Sarang Narkhede (2018). *Understanding Confusion Matrix*. Towards Data Science.
- [38] Boaz Smueli (2019). *Matthews Correlation Coefficient Is The Best Classification Metric You've Never Heard Of*. Towards Data Science.
- [39] OliverLegarreta. (2020, June 17). OliverLegarreta/TFG 1.2 (Version 1.2). Zenodo. <http://doi.org/10.5281/zenodo.3898899>
- [40] Bernal Toro, Francia Helena; Montoya Santacruz, Juan David. (2003). *ESTUDIO DE ESPECIES TOLERANTES A LOS INCENDIOS FORESTALES EN LA CUENCA MEDIA DEL RIO CALI*. Trabajo de grado para optar al título de Administrador del Medio Ambiente y de los Recursos Naturales. Universidad Autónoma de Occidente. Santiago de Cali.